Cellular and Molecular Life Sciences

REVIEW

Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence

Fernando Ruiz-Perez · James P. Nataro

Abstract Serine proteases exist in eukaryotic and prokaryotic organisms and have emerged during evolution as the most abundant and functionally diverse group. In Gram-negative bacteria, there is a growing family of high molecular weight serine proteases secreted to the external milieu by a fascinating and widely employed bacterial secretion mechanism, known as the autotransporter pathway. They were initially found in *Neisseria*, *Shigella*, and pathogenic *Escherichia coli*, but have now also been identified in *Citrobacter rodentium*, *Salmonella*, and *Edwardsiella* species. Here, we focus on proteins belonging to the serine protease autotransporter of Enterobacteriaceae (SPATEs) family. Recent findings regarding the predilection of serine proteases to host intracellular or extracellular protein-substrates involved in numerous biological functions, such as those implicated in cytoskeleton stability, autophagy or innate and adaptive immunity, have helped provide a better understanding of SPATEs' contributions in pathogenesis. Here, we discuss their classification, substrate specificity, and potential roles in pathogenesis.

Keywords SPATE · Autotransporter · Glycoprotein · Immune evasion · Cytotoxin

Abbreviations

AIDA-like	Adhesin involved in diffuse adherence-like
AIEC	Adherent-invasive *E. coli*
APEC	Avian pathogenic *E. coli*
DAEC	Diffuse-adhering *E. coli*
EAEC	Enteroaggregative *E. coli*
EHEC	Enterohemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ETEC	Enterotoxigenic *E. coli*
ExPEC	Extra-intestinal *E. coli*
HUS	Hemolytic uremic syndrome
REPEC	Rabbit pathogenic *E. coli*
SPATE	Serine protease autotransporters from Enterobacteriaceae
STEC	Shiga-toxin secreting *E. coli*
TA	Trimeric autotransporters
TPS	Two-partner secretion system
UPEC	Uropathogenic *E. coli*

F. Ruiz-Perez (✉) · J. P. Nataro
Department of Pediatrics, School of Medicine, University of
Virginia, P.O.Box 800326, MR4 Room 4012C, 409 Lane Road,
Charlottesville, VA 22908, USA
e-mail: fr3d@virginia.edu

Introduction

Pathogenic bacteria have emerged with highly specialized protein secretion systems. The type V secretion pathway is now considered the most common secretion mechanism used to deliver virulence factors by Gram-negative bacteria [1], which includes the classical monomeric autotransporters characterized by an N-terminal passenger and a C-terminal translocator domain (Va) [2]; the two-partner secretion system in which the passenger and the translocator domain harboring polypeptide-transport-associated (POTRA) motifs are encoded by separated genes (TPS,Vb) [3]; the trimeric autotransporters or Oca family, in which three passenger domains are fused to the one-third segment of a fully functional C-terminal translocator domain allowing secretion of trimeric polypeptides (Vc) [4]; the Vd family, exemplified by the PlpD autotransporter from *Pseudomonas aeruginosa*, in which passenger-POTRA-translocator domains are fused in a single polypeptide [5]; and the

Ve family, in which, unlike the classical autotransporters, the domain structure of their members is inverted, with an N-terminal translocator domain followed by a C-terminal passenger domain [6].

Among the classical monomeric autotransporters (Va), two distinctive groups can be recognized; those whose passenger domains are secreted to the extracellular milieu, and those whose passenger domains are exposed on the bacterial surface. The serine protease autotransporters from Enterobacteriaceae (SPATE) constitute a superfamily of virulence factors whose members resemble those belonging to the trypsin-like superfamily of serine proteases. They are generally secreted into the external milieu and are highly prevalent among enteropathogens, including *Shigella* and all *Escherichia coli* pathotypes [7, 8].

Several findings suggest that SPATEs have evolved with distinctive predilection to degrade host intracellular or extracellular substrates, which trigger a variety of adverse effects on host cells. Although SPATEs exhibit diverse amino acid sequences among cognates, they appear to elicit the same harmful effect on their host cells. Here, we discuss recent findings which have enlightened the mechanism of virulence and the potential roles in pathogenesis of this growing family of virulence factors.

The autotransporter pathway

In order to target their host cells, pathogenic Gram-negative bacteria secrete effector proteins into the periplasm, outer membrane, or external milieu by at least eight protein secretion pathways, designated types I to VIII [9]. The type V secretion system is the most widespread secretion pathway for protein transport across the outer membrane [1]. Classified in the type V secretion system are the autotransporter proteins (AT) [10, 11], whose denomination originated from the assumption (now apparently disproven) that all elements necessary for their translocation to the external milieu were encoded in the molecule itself. A number of recent studies have shown that autotransporters require accessory proteins located in the inner membrane [12, 13], periplasm [14–18], and outer membrane [12, 17–21] to reach their final destination: the bacterial surface or the extracellular milieu.

The general structure of the autotransporter proteins comprises three functionally different domains: the signal peptide, which targets the protein into the periplasm; the N-terminal passenger domain (also called the α-domain), which encodes the biological function of the AT-molecule; and the pore-forming C-terminal translocator domain (also known as the β-domain), which targets the protein to the outer membrane (OM) (Fig. 1). Although there are at least three hypothetical models to explain autotransporter

biogenesis [22–26], the evolving model [24, 27, 28] (Fig. 2) for autotransporter translocation comprises targeting of the AT-protein by its signal peptide into the periplasmic space in a Sec-dependent fashion, and which may occur co-translationally or following AT synthesis in the cytosol [29, 30]. Once in the periplasm, the AT proteins are protected and preserved in a "translocation-competent" state by common periplasmic chaperones, such as Skp, SurA, and DegP, toward the β-barrel assembly machinery (Bam) complex, which assists folding and insertion of outer membrane proteins (OMPs) into OM [14–18]. At this stage, the passenger domain commences its translocation through its C-terminal pore with further assistance of the Bam complex [17–21]. There is experimental evidence that translocation of the passenger domain across the channel begins with the C-terminus of the passenger followed by its distant N-terminus, until the entire passenger domain reaches the outside, a contortion known as the "hairpin model" [31, 32]. It was suggested that, during the translocation process, the passenger domain must be unfolded or partially folded in order to traverse the narrow AT pore, but completely folded once it reaches the bacterial surface [14, 32]. Also, since no ATP source exists in the periplasm, it was proposed that vectorial folding of the C-terminal passenger domain provides the necessary energy source for translocation and folding of AT proteins [31, 33, 34]. Once on the bacterial surface, the passenger domain may remain attached or get further incised from the AT translocator pore if a cleavage site exists between the alpha and beta domains (Figs. 1, 2). Passenger domains with adhesive properties, such as the autotransporter adhesins, often remain attached to the bacterial surface, but some passenger domains, such as those with protease activity, are generally released into the extracellular milieu. This latter class includes the SPATEs, the subject of this review.

Serine protease autotransporters of Enterobacteriaceae

The most abundant and functionally diverse proteolytic enzymes in living organisms are the serine proteases [35]. These enzymes cleave peptide bonds in protein substrates, in which a serine residue serves as the nucleophilic amino acid at the active site, thus accounting for their name. The active site of serine proteases is also dependent on two more residues, typically His and Asp, which, along with the Ser residue, form the catalytic triad. Serine proteases are grouped into two wide-ranging categories: the trypsin-like and subtilisin-like families. A more refined classification of these proteases can be found in the MEROPS database [36, 37], which classifies serine proteases into clans and families based on the catalytic mechanism and common ancestry, respectively [37].
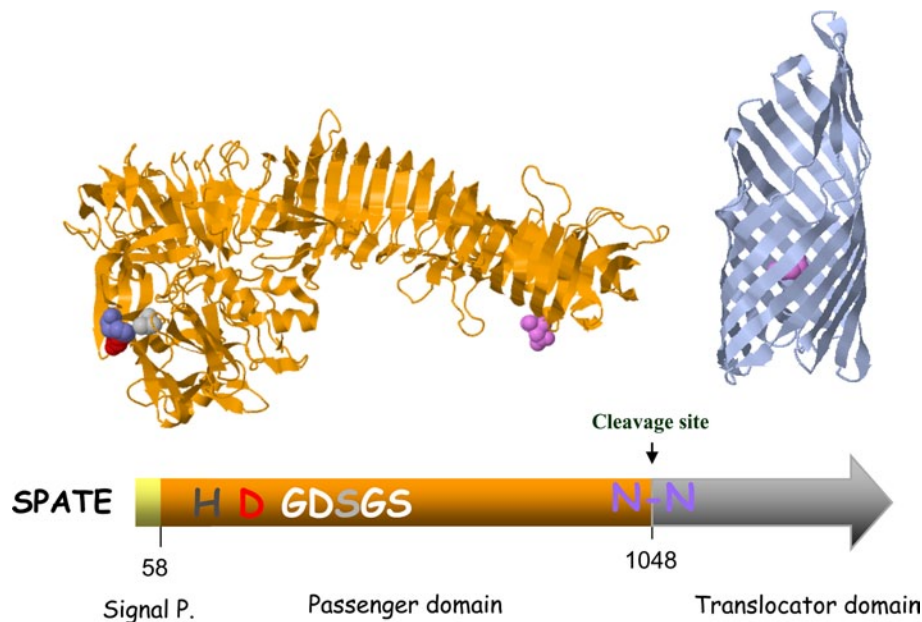
Fig. 1 General structure of the serine protease autotransporter from Enterobacteriaceae (SPATE). Autotransporters comprise three functionally different domains: the signal peptide, which targets the autotransporter to the inner membrane SecYEG translocon; the N-terminal passenger domain (also called α-domain), which encodes the biological function of the AT-molecule; and the pore-forming C-terminal translocator domain (also known as the β-domain). Here, the crystal structure of a prototype SPATE protease; EspP passenger and translocator domains are depicted (PBD ID:3SZE and PBD ID:3SLT). The passenger domain (*orange*) shows the characteristic serine protease motif GDSGS on SPATE proteins along with the residues involved in the formation of the catalytic triad (His, Asp and Ser). The passenger domain is incised between the two Asn residues (*violet*) bridging passenger and translocator domains inside the pore during the passenger secretion. Residues in the catalytic triad and cleavage site between passenger and translocator domains are illustrated in *colored letters* and *spheres*

In prokaryotes, the serine proteases are involved in numerous biological processes, such as those associated with metabolism, development, and virulence. In Gram-negative bacteria, the majority of serine proteases secreted by the autotransporter pathway are implicated in virulence [1]. In this rubric, SPATEs constitute a superfamily of virulence factors whose members resemble those belonging to the trypsin-like superfamily of serine proteases [8, 38]. SPATE proteins are produced by enteric pathogens including *E. coli*, *Shigella*, *Salmonella*, *Edwardsiella*, and *Citrobacter* species. Interestingly, SPATEs have been found not only in all recognized *E. coli* pathotypes including enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC), enterotoxigenic E. coli (ETEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC), and enteroaggregative *E. coli* (EAEC), all agents of enteric/diarrheal disease, but also in extraintestinal *E. coli* pathogens (ExPEC), such as uropathogenic *E. coli* (UPEC) and septicemic *E. coli*, which are responsible for urinary tract infections (UTIs) and sepsis/meningitis, respectively.

With the major progress in bacterial genome sequencing, new and previously described SPATE proteins are being uncovered in other human and animal pathogens, including *Salmonella*, *Citrobacter*, and *Edwardsiella*, and less frequently in commensal strains.

Structural organization of SPATEs

The term SPATE was proposed more than a decade ago to describe serine proteases produced by the Enterobacteriaceae family with a similar architecture and functional motifs, including the presence of a long signal peptide (48–59 aa); the presence of a catalytic triad formed by H, D, and S residues, in which the motif GD**S**GS harbors the catalytic serine; the existence of two consecutive asparagines (N–N), which link the passenger domain to a translocator domain; and the presence of very conserved 5–10 residue motifs on passenger and beta domains, whose significance is not yet fully understood (reviewed in detail by Yen et.al. [8]). With few exceptions [39], the mature SPATE protein or the passenger domain is processed inside the pore by an unusual autocatalytic reaction [40–42], and released into the extracellular space (Fig. 1).

The first solution structure of a SPATE passenger domain was that of hemoglobin binding protein (Hbp) from *E. coli* EB1 [43], a strain isolated from an intra-abdominal
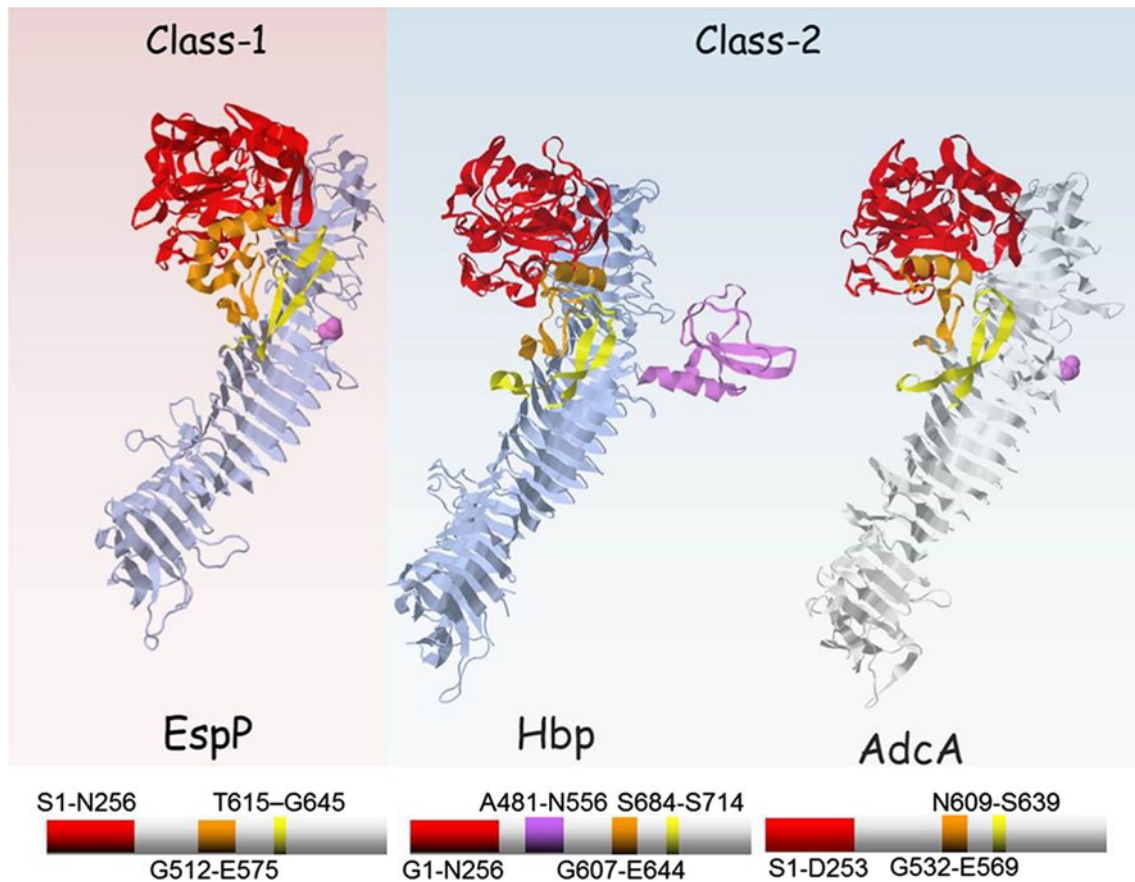
Fig. 2 Prevailing model of AT biogenesis. The AT molecule is targeted to the periplasmic space by the Sec apparatus. Once in the periplasm, the AT intermediate is stabilized by periplasmic chaperones such as Skp, SurAm and/or DegP or, when a chaperone shortage exists, perhaps by FkpA. This step is believed to prevent non-productive aggregation, premature folding and/or to maintain the species in a partially folded "translocation-competent state". Recent data suggest that the Bam complex catalyzes both the integration of the β-domain into the OM, and the translocation of the passenger domain across the OM in a C- to N-terminal direction, followed by intra-barrel cleavage and disengagement of the passenger domain (in the case of SPATEs)

wound infection. The Hbp structure revealed an architecture rich in β-strands arranged in a parallel β-helical structure (Fig. 3). The fold was similar to that observed in the non-SPATE pertactin from *Bordetella pertussis* [44], and subsequently also shown by other autotransporters, including immunoglobulin A1 protease from *Haemophilus influenzae* [45]; adhesion and penetration protein (Hap) from *H. influenzae* [46]; and the vacuolating cytotoxin autotransporter (VacA) from *Helicobacter pylori* [47]. With the exception of the EstA esterase from *P. aeruginosa*, whose passenger domain revealed no β-strands but α-helices and loops [48, 49], a β-stranded architecture has also been predicted in 97 % of 507 other AT-passenger sequences using bioinformatic analysis [34], strongly suggesting that most AT passenger domains adopt a right-handed parallel β-helix structure [34]. Initially, it was proposed that, in Pertactin-a non-SPATE AT, the β-helical stalk might have a structural role in the translocation of the passenger domain across the bacterial outer membrane, by stabilizing the passenger in

the extracellular space following its translocation [50]. A region in the β-helical stalk termed auto-chaperone (AC) was anticipated to assist folding of the passenger simultaneously with, or following translocation across, the outer membrane [50]. Subsequently, it was shown that the β-helix stalk folded reversibly in isolation, and was not prone to aggregation during folding, which led authors to propose that the high stability observed in the C-terminal rungs of the beta-helix might serve as a template for the formation of native protein during OM secretion, and that vectorial folding of the beta-helix could contribute to the energy-independent translocation mechanism, largely hypothesized for AT translocation [34]. Similarly, it was shown that efficient secretion of EspP passenger domain requires the stable folding of a C-terminal 17-kDa segment in the β-helical stalk [32]. Mutations that perturb the folding of this segment did not affect its translocation across the OM but impaired the secretion of the remainder of the passenger domain. Hence, it was suggested that translocation of

Fig. 3 Stereo ribbon diagrams showing the overall structure of the passenger domain of class 1 and class-2 SPATEs. EspP(3SZE) and Hbp(1WXR) PDB annotations were used to model prototype structures of class-1 and class-2 SPATE proteins, respectively. By using the SWISS-MODEL [162], Jmol (http://www.jmol.org/) and SignalP 4.1 servers [163], along with the two SPATE PDB annotations, the hypothetical structure of AdcA; a class-2 SPATE lacking of domain-2 was modeled (QMEAN4 score 0.639, *z* score:−2.02). Helices and strands are colored *red* for the protease domain (Domain-1; residue S1–N256 in EspP, G1–N256 in Hbp and S1–D253 in AdcA). Domain-2 is shown in *violet*; predicted as a chitinase-like domain in Hbp (residues A481–N556), but not present in EspP and AdcA. Domain-3, which forms a helix–turn–helix motif facing domain-1 is shown in *orange* (residues G512–E575 in EspP, residues G607-E644 in HbP, and G532-E569 in AdcA). Domain-4 is shown in *yellow* (residues T615–G645 in EspP, S684-S714 in Hbp, and N609-S639 in AdcA). The β-strands in the parallel β-helix, classical in most AT proteins, are colored *gray* for EspP and Hbp, but *white* for the hypothetical AdcA structure. *Bars* underneath the structures illustrate positions of globular regions in the mature passenger domain

the entire EspP passenger domain proceeds through a vectorial diffusion and folding process nucleated by the formation of a C-terminal core, and that the repetitive β-helical architecture of passenger domains may have evolved to promote the sequential movement of a polypeptide in an environment devoid of ATP [32, 33]. The role of the passenger C-terminal β-helical segment in folding and secretion of the residual AT-passenger was challenged when the solved crystal structure of the EstA AT protein lacking the β-helical stalk subdomain was published, which is fully functional in exporting is native globular passenger domain onto the extracellular surface, despite no β-helical structure being within it [48]. Consequently, this finding suggested that the β-helical stalk may have an additional, and perhaps more specialized, role in ATs. For instance, it was shown

that the EspP and EspC SPATEs oligomerize to form coiled-coil megastructures that possess adhesive and cytopathic activities on host epithelial cells [51]. EspP monomers aggregated in a concentration-dependent manner to form "rope-like" structures of stable, highly organized β-sheet structures, which showed striking biochemical and biophysical similarities to human neurodegenerative and amyloidal-related diseases such as Alzheimer, Parkinson, and prion diseases [51]. Multimerization of ATs mediated by the β-helical stalk has also been shown in a family of self-associating autotransporter proteins (SAATs), whose members are involved in adhesion and biofilm formation [46, 52]. Crystal structure of the passenger domain of the prototype SAAT, *H. influenzae* Hap, revealed a β-helical C-terminal triangular prism-like structure that appears to be
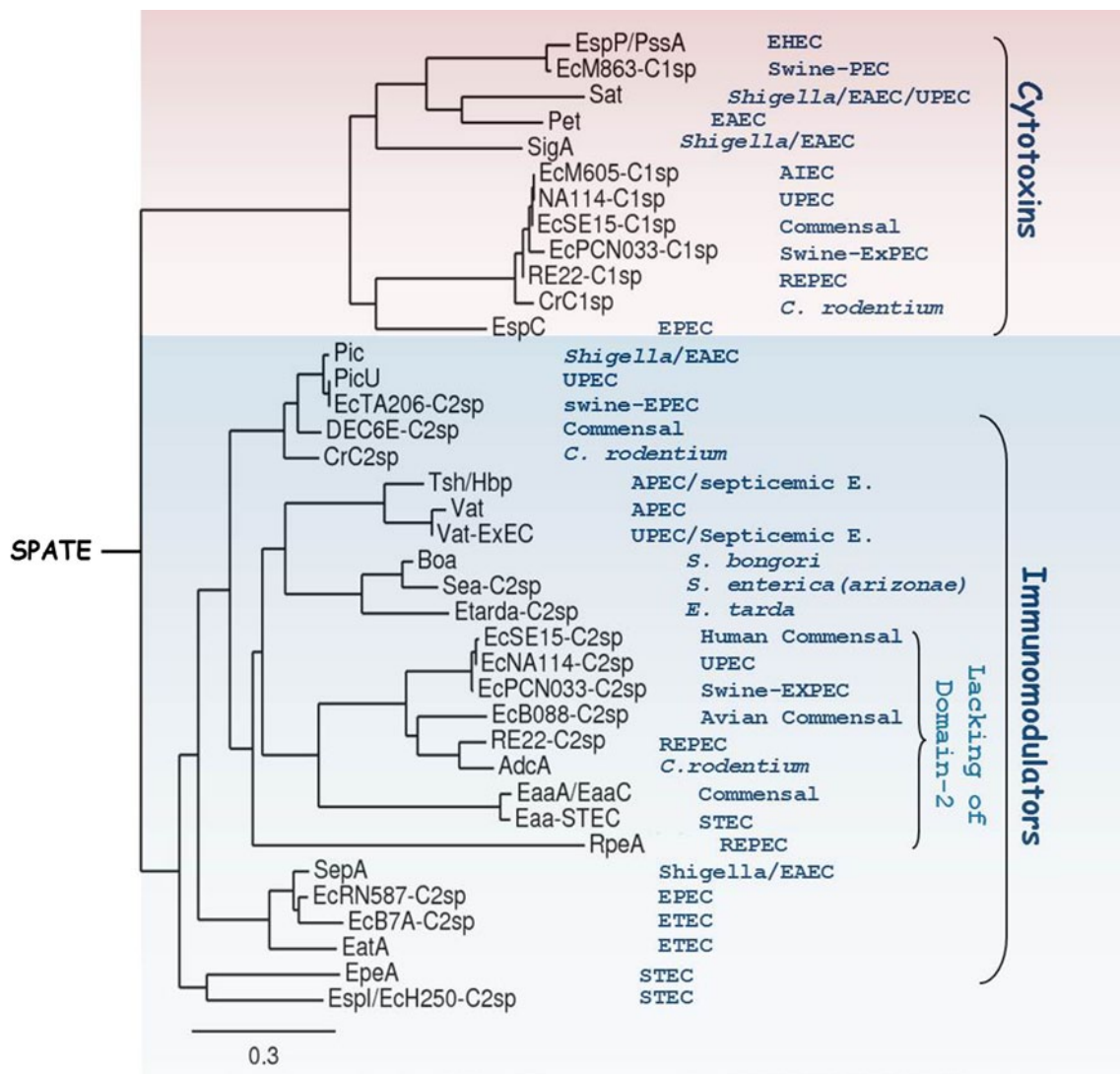
conserved among SAAT autotransporters, and that mediates formation of Hap–Hap dimmers into higher-order oligomerization through the β-helical stalk, suggesting the mechanistic principle for bacterial aggregation and biofilm formation mediated by SAAT [46]. Hence, taken together, these findings suggest that, when present, the β-helical stalk may serve as mediator of binding to host organisms.

Beside the β-helical stalk, the Hbp SPATE solution structure also showed the presence of two N-terminal globular domains, termed domain-1 and 2 (Fig. 3). Domain-1 contains the catalytic triad of protease consisting of residues 1–256 of the Hbp mature protein, while domain-2 comprises residues 481–556, with an apparent resemblance to the chitin-binding region of chitinase, and which was

Fig. 4 Domain-2 and -3 among SPATE autotransporters proteins. Alignment of the aminoacid sequence of 28 SPATE passenger domains with Clustal-Omega [164] revealed differences in domain-2 and -3 between SPATE classes. Class-1 SPATEs (*pink*) lack domain-2, while it is present in only a number of class-2 SPATEs (*blue*). Domain-3 in class-1 SPATEs is larger than its equivalent in class-2 SPATEs and includes a potential disulphide bond (two cysteines), which is missing in class-2 SPATEs. Identical residues are shaded in *black*, similar residues are shaded in *gray*, conserved cysteines(C) in domain-3 class-1 SPATEs are shaded in *blue*. Aminoacid sequence of Tsh/Hbp domain-2(A481-N556), and the aminoacid sequences of domain-3 for EspP(G512-E575), AdcA(G532-E569), and Tsh/Hbp(G607-E644) are shown in *blue* and correspond to the globular regions illustrated in Fig 3

Fig. 5 Sequence relationships of SPATE proteins. Phylogenetic analysis of the aminoacidic sequence of the SPATE passenger domain reveals two distinctive classes of SPATE proteins: denominated class-1, cytotoxic; and class-2, lectin-like immunomodulators. Class-2 spates also includes a cluster of SPATEs found mostly in animal pathogens, which lack the classic domain-2 (see also Figs. 3, 4). Within both classes, clusters of allelic variants are also appreciated.

For instance, allelic variants for EspP, Cr-C1sp, Pic, Tsh/Hbp, Boa, EcSE15-C2sp, EaaA/EaaC, SepA, and EpeA are observed. Bacterial species from which SPATE sequences were originated are shown on the *right side*. Phylogenetic analysis was performed as follows: alignment of sequences was built with Clustal-Omega [164] and cured by Gblocks [165]. Phylogenetic tree was constructed with PhyML/bootstrapping [166] procedure and visualized with TreeDyn [167]

believed to play a role in substrate recognition, though this assumption remains controversial, given that domain-2 is not present in all SPATEs and that removing this domain from the Hbp molecule did not affect its proteolytic activity on known substrates [43, 53].

More recently, the structure of the passenger domain of EspP (extracellular serine protease plasmid-encoded) produced by EHEC O157:H7 was elucidated [54]. Like the Hbp protein, the full-length passenger of the EspP protein comprised structural subdomains: a globular subdomain shaped by residues 1–256 and corresponding to domain-1

in Hbp, two more globular domains termed domain-3 and -4, and a C-terminal β-helical stalk domain (Fig. 3). As expected, the EspP domain-1 displayed high similarity to the folds of the Hbp passenger domain-1, which resembles those seen in the chymotrypsin family of serine proteases (Fig. 3). Unlike Hbp protease, the EspP protein does not contain the chitinase-like domain-2 protruding from the β-helical spine. Instead, a single residue in EspP substituted the domain-2 formed by 76-residues in Hbp (Figs. 3, 4). Furthermore, the equivalents of the domain-3 and -4 of EspP are also present in Hbp (Figs. 3, 4). Interestingly,

the EspP subdomain-3 forms a helix–turn–helix motif that rests on one face of the domain-1, and a disordered loop in which the only two cysteine residues in the entire EspP passenger domain reside. Similarly, the only two cysteines of Pet autotransporter were found to lie in the region equivalent to the domain-3 in Pet, termed domain2A, and this seems to be a characteristic shared by many other SPATEs, but not observed in Hbp/Tsh (Fig. 4) [55].

The only two solved structures of SPATEs have assisted in structural and functional studies on all other SPATEs whose protein structures are not yet available.

Classification of SPATEs

The SPATE family, which now includes more than 25 proteases with apparent diverse substrates, have been phylogenetically divided into two distinct classes based on the amino acid sequence of the passenger domain [2, 8]. In Fig. 5, we show the phylogenetic tree analysis of all SPATEs whose amino acid sequences are to date available. The bifurcation of SPATEs into class-1 and class-2 is consistent with previous phylogenetic trees generated using split decomposition analysis [8, 56], where beside the amino acidic discrepancy, members of class-1 also lack the domain-2 shown by the solved structure of the Hbp protease, whereas all previously known class-2 SPATEs retained this domain (Figs. 3, 4).

Recently, a SPATE group from newly sequenced genomes showed high homology to class-2 SPATEs, but they lack domain-2 (Figs. 3–5). This domain2-less class-2 subgroup also draws attention to the fact that they are mostly produced by animal pathogens including: the AdcA protein produced by the mouse pathogen *C. rodentium* [57] (Fig. 3); the RpeA protein from the rabbit *pathogenic E. coli* (REPEC) [39]; a new protease here termed RE22-C2sp (RE22 class-2 serine protease, NCBI Accession No. ZP_03046500.1) also from REPEC [58]; a new protease here termed EcPCN033-C2sp (*E. coli* PCN033 class-2 serine protease, NCBI Accession No. EGP21815.1) from a swine pathogenic *E. coli* [59]; and the protein here termed EcB088-C2sp (*E. coli* B088 class-2 serine protease, NCBL Accession gi/29342089) produced by an avian *E. coli* strain. Employing the automated protein structure homology-modeling server SWISS-MODEL [60] and the open-source Java viewer for chemical structures in 3D, Jmol (http://www.jmol.org/), along with the crystal structures of Hbp [43] and EspP [54], we modeled the stereo ribbon diagram for AdcA, the only protein experimentally studied so far belonging to the domain-2-less class-2 SPATE cluster, and which was previously mistakenly compared to the AIDA family of surface-exposed autotransporter proteins [57] when it actually belongs to the SPATE family

(Figs. 3, 4, 5). Along with the absence of the domain-2, the hypothetical structure of AdcA revealed high similarity to globular regions of class-2 compared with class-1 SPATEs, including the discrete domain-3 facing the proteolytic domain-1 (38 residues in AdcA and Hbp), while the equivalent region in EspP is more voluminous (64 amino acid residues in size) [43, 54] (Figs. 3, 4).

The variable presence of domain-2 among class-2 SPATEs is enigmatic, and its potential involvement in substrate recognition is still unclear [43, 53]. Removal of the domain-2 from Hbp did not abolish protease activity on synthetic peptides, which are known to be cleaved by Hbp [53, 56]. Similarly, domain-2 expressed as a separate polypeptide did not bind to Tsh/Hbp substrates such as the heme group, fibronectin, and collagen-IV [53]. Consequently, more refined studies including on O-glycoproteins, the natural substrates of this subset of SPATEs [61], should be carried out to properly address the actual role of the domain-2.

The fact that the domain-2 is not found in all class-2 SPATE members makes this domain no longer a classifying feature of class-2 SPATEs. Instead, a discrete domain-3 lacking cysteines seems to be a characteristic of class-2 SPATEs, while a bulky domain-3 harboring two contiguous cysteines with a potential to form a disulfide bond is more distinctive of class-I SPATEs [54, 55] (Figs. 3, 4). Since domain-3 faces one side of the domain-1 [54, 55] (Fig. 3), where substrates are incorporated and cleaved, it is tempting to hypothesize that the domain-3 alongside the domain-1 may have a certain role in substrate selectivity. In an early study, transposon-based scanning linker mutagenesis on the cytopathic serine protease Pet identified a subset of insertions encompassing the region equivalent to domain-3 and regions flanking this domain [62] with proteolytic but not longer cytopathic effects. Those mutants displayed decreased binding and internalization upon incubation with HEp-2 cells, which made the authors postulate the involvement of this region in cytotoxicity [62]. Close proximity of Pet domain-3 to domain-1 was recently shown from a three-dimensional model for Pet built using the crystal structure of the Hbp protein [55]. The model suggested that the equivalent region for domain-3 in Pet (termed domain 2A) contacts domain-1 and returns to the β-helical stem close to the point of departure [55]. However, whether the domain-3 interacts with substrates and gives specificity to SPATEs remains as yet unknown.

Another possible role for domain-3 is to restrict the accessibility of substrates to domain-1. Class-2 SPATEs such as Pic, PicU, Tsh/Hbp, and EpeA, for example, are able to cleave huge substrates such as the heavily glycosylated O-linked glycoproteins (the mucins), which would be the result of lesser impedance of substrate access to domain-1's groove by the small globular domain-3, while

Table 1 Serine protease autotransporter from Enterobacteriacea

Protein	Species	Disease	Function/effect	Biological substrate	Ref.
Class-1 SPATEs					
EspP/PssA EcM863-C1sp	EHEC Swine-EPEC	Bloody diarrhea, hemorrhagic colitis/HUS	Cytotoxin, adherence, biofilm	Factor V C3/C3b,C5	[88, 92, 93, 99, 105, 109–111]
Sat	UPEC *Shigella* EAEC, DAEC	Cystitis, pyelonephritis Bloody, mucoid diarrhea Watery diarrhea	Cytotoxin, enterotoxin, impairment of tight junctions, autophagy	α spectrin Factor V	[71, 91, 127, 129–132]
Pet	EAEC	Watery diarrhea	Cytotoxin Enterotoxin	α spectrin FAK Factor V	[63, 66, 89, 102, 113–117]
SigA	*Shigella*, EAEC	Bloody, mucoid diarrhea Watery diarrhea	Cytotoxin Enterotoxin	α spectrin	[64, 71, 72, 90, 123]
EspC	EPEC	Watery diarrhea	Cytotoxin, Enterotoxin	α spectrin Factor V haemo- globin	[56, 65, 95, 118–121]
EspC-like: CrC1sp RE22-C1sp EcPCN033-C1sp EcNA114-C1sp EcM605-C1sp	*C. rodentium* REPEC Swine-ExPEC UPEC AIEC	Diarrhea Diarrhea Meningitis, Septicemia Prostatitis Crohn's Disease	Unknown EspC-like effects?	Unknown EspC-like sub- strates?	See text
Class-2 SPATEs					
Pic	*Shigella* EAEC	Bloody, mucoid diarrhea Watery Diarrhea	Mucinase Immunomodulation Colonization Lectin-like hemagglutinin Mucus secretion activity Serum resistant	Mucin -O-linked glycopro- teins: CD43, CD44,CD45, CD93, CD162,CX3CL1	[56, 61, 68, 69, 104, 141]
Pic-like: PicU EcTA206-C2sp	UPEC Swine-EPEC	Cystitis, pyelonephritis Diarrhea	Mucinase Pic-like effects?	Mucin Pic-like substrates?	[75, 76] See text.
Cr-C2sp	*C. rodentium*	Diarrhea	Unknown Pic-like effects?	Pic-like substrates?	See text.
Tsh/Hbp	Septic *E. coli* APEC	Wound Infections Septicemia	Mucinase Immunomodulation Iron sequestering Lectin -like hemagglu- tinin ECM proteins binding	Mucin Hemoglobin O-linked glyco- proteins: CD43, CD45,CD93, CD162, CXCL1	[61, 78, 79, 134, 135, 137–139]
Vat	APEC	Septicemia, respiratory disease, cellulitis	Vacuolating agent Tsh/Hbp-like effects?	Unknown Tsh/Hbp-like sub- strates?	[76, 80, 140]
Vat-like: Vat-ExEc	UPEC Septicemic *E. coli*	Cystitis, pyelonephritis Septicemia, meningitis	Tsh/Hbp-like effects?	Unknown Tsh/Hbp-like sub- strates?	[76] See text.
Boa *Boa like:* Sea-C2sp Etarda-sp	*Salmonella* Salmonella Edwardsiella	Diarrhea Diarrhea Septicemia	Tsh/Hbp-like effects?	Unknown Tsh/Hbp-like sub- strates?	See text
AdcA *AdcA-like:* RE22-C2-sp EcNA114-C2-sp EcPCN033-C2- sp Eaa-STEC RpeA	*C. rodentium* REPEC UPEC Swine ExPEC STEC REPEC	Diarrhea Diarrhea Prostatitis Meningitis, septicemia Diarrhea/HUS Diarrhea	Adherence Tsh/Hbp-like effects? Intestinal colonization	Unknown Tsh/Hbp-like sub- strates?	[39, 57] See text

Table 1 Continued

Protein	Species	Disease	Function/effect	Biological substrate	Ref.
SepA EcRN587-C2sp	*Shigella* EAEC EPEC	Bloody, mucoid diarrhea Watery diarrhea	Enterotoxin SepA-like effects?	Unknown	[56, 61, 71, 136, 156–159] See text.
EatA *EatA-like:* EcB7A-C2sp	ETEC	Watery diarrhea	Enterotoxin EatA-like effects?	EtpA adhesin EatA-like substrates?	[87, 160] See text.
EpeA Espl	STEC	Diarrhea, HUS	Mucinase Pic-like effects?	Mucin Pic-like substrates?	[70, 88]

SPATE serine protease autotransporters from enterobacteriaceae, *HUS* hemolytic uremic syndrome, ? predicted activity by homology, *ECM* extracellular matrix, *EPEC* enteropathogenic *E. coli*, *EHEC* enterohemorrhagic *E. coli*, *EAEC* enteroaggregative *E. coli*, *ETEC* enterotoxigenic *E. coli*, *STEC* shiga toxin-producing *E. coli*, *UPEC* uropathogenic *E. coli*, *AIEC* adherent invasive *E. coli*, *ExPEC* extra intestinal *E. coli*, *APEC* avian pathogenic *E. coli*

the much larger disulphide bond-containing domain-3 of class-1 SPATEs may dodge the accessibility for huge proteins. Indeed, substrate specificity is a distinctive characteristic of SPATEs; class-1 SPATEs are cytotoxic in vitro and cause mucosal damage in intestinal explants [63–67] (Table 1). Class-1 SPATEs are therefore believed to degrade primarily intracellular substrates. On the other hand, most class-2 SPATEs cleave mucin and exhibit a subtle competitive advantage in mucosal colonization [56, 63, 68–70], but they have no obvious predilection for any intracellular substrates; consequently, class-2 SPATEs are believed to mainly target extracellular substrates (Table 1). In fact, we have recently found that class-2 SPATEs target a broad range of extracellular glycoproteins present not only on the epithelial cells, which are lining the intestinal mucosa, but also on the surface of nearly all lineages of hematopoietic cells [61].

Allelic variation among SPATEs

Interestingly, pathogens often harbor more than one SPATE protein, typically a combination of at least one SPATE from each class. *Shigella flexneri*, for instance, almost invariably carries SigA (class-1), and many strains also carry SepA and Pic (class-2); UPEC carries PicU and Vat-ExEc (class-2), and Sat (class-1); the prototype EAEC strain 042 carries Pet (class-1) and Pic (class-2); but we have found that the vast majority of EAEC strains possess at least up to 3 SPATE genes including SepA, Pic, SigA, and Sat [71, 72]. Combinations of class-1 and 2 SPATEs can also be seen in animal pathogens including the mouse pathogen *C. rodentium*, rabbit enteropathogenic *E. coli*, and swine pathogenic *E. coli* (Fig. 5; Table 1). The significance of highly prevalent SPATEs and their occurrence in multiple numbers in a single pathogen are under study.

Identity/similarity analysis of the amino acid sequence of SPATE passenger domains using the Matrix Global Alignment Tool (MatGAT) [73] showed approximately 28–34 % amino acid identity among class-1 and class-2 SPATEs. However, members of the same class share roughly 45–75 % amino acid identity. Among all the SPATE sequences described so far, class-2 members outnumber those belonging to class-1, mainly because of the high occurrence of allelic variation among their members, which share identity/similarities from 87.7/93.1 to 96.5/97 %. A clear example of allelic variation is observed among Pic-homologues; protease involved in colonization (Pic) protein is produced by *Shigella flexneri* 2a and several EAEC strains, including the deadly German outbreak enteroaggregative *E. coli* O104:H4 strain [69, 71, 72, 74]. Pic has 95.7 % amino acid identity to PicU found in the extra-intestinal UPEC pathogen [75, 76]. PicU is also found in the *E. coli* strain 83972, originally isolated from a young girl who had carried it for at least 3 years in an asymptomatic bacteriuria (ABU) [77], and in the recently published swine diarrheagenic *E. coli* TA206 genome (here termed EcTA206-C2sp for *E. coli* TA206 class-2 serine protease; NCBL Accession no.: gi/331656291).

Interestingly, with lesser but still significant homology to Pic are two hypothetical SPATE proteins identified in the commensal *E. coli* DEC6E genome, here designated DEC6E-C2sp (DEC6E class-2 serine protease; NCBI Accession no. EHV62301.1), and in *C. rodentium*, here termed Cr-C2sp (*C. rodentium* class-2 serine protease; NCBI Accession no.: YP_003368482.1), which share ~80 and 74.6 % amino acid identity with Pic, respectively.

Similar to the cluster of Pic variants, at least 3 other clusters among class-2 SPATEs can be appreciated from the phylogenetic tree (Fig. 5), suggesting that each SPATE cluster originated from allelic variation of a single parental SPATE. For instance, the Tsh protein from avian pathogenic *E. coli* (APEC) [78], which differs in only 2 residues with the Hbp protein from human pathogenic *E. coli* EB1 [79], shares 73.2 % amino acid identity with the Vat

protease from APEC [80], and the latter is 96.5 % identical to a class-2 SPATE protein indistinctly annotated as hemoglobin protease or temperature-sensitive hemagglutinin, and highly prevalent in extra-intestinal pathogenic *E. coli* (ExPEC) strains such as UPEC and septicemic *E. coli* [81, 82], and here, for a simplified designation and given that it is closer to Vat (96.5 % identity) than to Tsh/Hbp (69.6 % identity), we have termed it Vat-ExPEC protease (Vat from extraintestinal pathogenic *E. coli*). Closer to Tsh/Hbp are also two SPATEs found in the Salmonella species, Boa from *S. bongori* [8, 83] and Sea-C2sp [*Salmonella enterica* subsp. arizonae class-2 serine protease, NCBI Accession no. YP_001570771.1) from *S. enterica* ssp. arizonae, both sharing 87.7 % amino acid identity with each other, but 50.1–51.8 % amino acid identity with Tsh. Interestingly, the agent causing septicemia in fishes and an opportunist pathogen in immune-compromised humans, *Edwardsiella tarda*, possesses a SPATE protein designed here, Edtarda-C2sp (*Edwardsiella tarda* class-2 *s*erine protease, NCBL Accession no.: ZP_06713707.1), which shares 50.1 % amino acid identity with the Tsh/Hbp protein, produced by strains known to cause septicemia in birds and humans [78, 79].

An interesting cluster of class-2 SPATEs is the aforementioned group of SPATEs lacking the domain-2; the characteristic N-terminal protein protuberance exhibited by Hbp and many other class-2 SPATEs. The domain-2-less cluster is more similar to the Tsh/Hbp cluster than the clusters grouping the Pic or SepA proteases, and this cluster encloses allelic variants mainly produced by animal pathogens including the hypothetical EcPCN033-C2sp protein encoded by the genome of the pathogenic *E. coli* PCN033 strain, originally isolated from the brain of a convalescent swine [59]; its allelic variant EcNA114-C2sp encoded by the multidrug resistance UPEC NA114 strain, originally isolated from the urine of a 70-year-old male patient with prostatitis in India [84] (Fig. 5); and several SPATEs found in commensal strains such as the EaaA/EaaC protease from *E. coli* [85] and hypothetical proteases with NCBI Accession nos.: BAI57550.1, here termed EcSE15-C2sp (*E. coli* SE15 class-2 SPATE protease) and ZP06661330.1, here termed EcB088-C2sp (*E. coli* EB088 class-2 SPATE) encoded by the genome of the human commensal *E. coli* SE15 (O150:H5) [86], and the avian commensal *E. coli* EB088, respectively. Two domain-2-less SPATE members have been experimentally studied so far; the AdcA protease produced by *C. rodentium* and RpeA from REPEC [39, 57]. AdcA shares 80.3/86.5 % identity/similarity with a hypothetical protein found in REPEC (NCBI reference sequence: ZP_03046499.1), here denominated RE22-C2sp. While RpeA, the only non-secreted protease from the SPATE family, shares only approximately 37–40 % identity with any member of the domain-2-less SPATE cluster and to Tsh [39].

The last cluster of class-2 SPATEs is typified by SepA produced by *Shigella flexneri* and EAEC, which shares 81/89 % identity/similarity with two almost identical hypothetical SPATE proteins present in EPEC and ETEC, here denominated ERN587-sp (*E. coli* ERN587 serine protease; Accession no.: EFZ76879.1) and B7A-sp (*E. coli* B7A serine protease, Accession no.: ZP_03031082.1), respectively. With lesser, but significant homology to SepA (roughly 71/83 % identity/similarity) is the previously described EatA from a typical ETEC H10407 strain [87]. EatA, in turn, shows 63–69 % similarity to EpeA from EHEC [70] and EspI from STEC [88]; the latter has an allelic variant of the STEC EH250 strain (98 % similarity, Accession no.: EGW90721.1) (Fig. 5).

The other branch of the SPATE phylogenetic tree embodies members of the class-1 category (Fig. 5). Two major clusters can be observed: one containing the classical SPATEs found in *Shigella* and EAEC such as Pet [89], SigA [90] and Sat [91], and the SPATE found in EHEC isolates termed EspP [92], also known as PssA [93]. An allelic variant of EspP is found in the swine pathogenic *E. coli* M863 genome, here denoted EcM863-C1sp (*E. coli* M863 class-1 serine protease; NCBI Accession no.: EGB59948.1), which shares 89/94 % identity/similarity with EspP. Sat and Pet are the two class-1 SPATEs showing higher identity/similarity (53.8/71.1 %) than any other member of their class, which may explain the comparable protease strength on their biological substrate, α-spectrin [67, 94].

SigA has roughly 45/68 % identity/similarity with Pet and Sat, while EspC, a protease encoded in a pathogenicity island of EPEC isolates [95], shows high homology with newly identified hypothetical SPATE proteins encoded in the genome of human and animal pathogens, such as one of the three SPATEs found in the mouse pathogen *C. rodentium*, here denominated CrC1sp (*C. rodentium* class-1 serine protease; NCBI Accession no.: YP_003368469.1), which shows 55.1/72.4 % identity/similarity to EspC. An allelic variant of Cr-C1sp is found in the rabbit pathogenic *E. coli* E22 strain, here termed RE22-C1sp (REPEC E22 class-1 serine protease, NCBI Accession no.: ZP_03046499.1), which shares 51.5/68.8 % identity/similarity with EspC and 85.3/89.8 identity/similarity with Cr-C1sp. Many other allelic variants of E22-C1sp (93–97 %, similarity) are hypothetical SPATE proteins with NCBI Accession nos.: EGP21815.1 (here termed EcPCN033-C1sp, *E. coli* PCN033 class-1 serine protease), ZP_08351236.1 (here termed EcM605-sp, *E. coli* M605 serine protease), BAI57554.1 (termed EcSE15-C1sp, *E. coli* SE15 class-1 serine protease) and AEG39156.1 (termed EcNA114-C1sp, *E. coli* NA114 class-1 serine protease). The PCN033 strain is a swine extra-intestinal pathogenic *E. coli*, the HM605 strain is a Crohn's Disease-associated adherent-invasive *E. coli* (AIEC) [96], the SE15

strain is a commensal strain [86] and the *E. coli* NA114 is a multidrug resistant UPEC strain isolated in India [84]. The high homology among these SPATEs and EspC suggests that all were derived from an EspC-like ancestor.

Evolution of SPATEs

Phylogenetic analysis of the SPATE passenger domain revealed two distinctive SPATE classes [8, 56]. In Fig. 5, a cluster of SPATEs with their respective allelic variants are shown. We observed that the extent of variation in some allelic members with respect to their cognates in the same cluster is more accentuated, which gives the impression that they have evolved to originate a distinct SPATE protein, such as in the case of EspC and CrC1sp, Pic and CrC2sp, Tsh and Vat, Boa and Etarda-C2sp, AdcA and EcNA114-C2sp, and SepA and EatA. Indeed, phylogenetic analyses performed on 120 members of the AT family derived from 20 bacterial genera [97] suggested that all members of the AT family have arisen almost exclusively by speciation and late gene duplication events within a single organism, while horizontal transfer of genes encoding ATs between distant organisms was a rare evolutionary event [97]. Certainly, we can envision that these two events may have occurred on the SPATE evolutionary tree. The large number of allelic variants among SPATEs, as well as the presence of two or more genes encoding SPATEs from one or both classes in *Shigella* (Pic, SepA, and SigA), EAEC (Pic, Sat, SepA, SigA and Pet), UPEC (Tsh, PicU and Sat), STEC (EpeA, EspP and EspI), REPEC (RpeA, E22-C1sp and E22-C2sp), *C. rodentium* (Cr-C1sp, Cr-C2sp and AdcA), swine ExPEC (EcPCN033-C1sp and EcPCN033-C2sp), and other pathogens, supports the speciation hypothesis, while a rare duplication event may have occurred in the *E. coli* ECOR-9 strain, which harbors two identical genes, *eaaA* and *eaaC* [85], and which may have also occurred in other non-SPATE ATs, such as the three copies of the Ag43 autotransporter found in EAEC [98]. Additionally, it was suggested that horizontal gene transfer of the SPATE Boa may have occurred within *Salmonella bongori* [8] because of the absence of SPATEs in all other *Salmonella* spp. and its close resemblance to *E. coli* SPATEs. It is not clear if Boa was an isolated event of lateral transfer, or is a common SPATE among these species given that its allelic variant Sea-C2sp is present in another *Salmonella* species (*S. enterica* ssp. *arizonae*), while its close homologue Sea-C2sp is found in *Edwardsiella tarda*—all human pathogens but common inhabitants of the intestinal tract of reptiles (Fig. 5). Nevertheless, occurrence of horizontal transfer of SPATEs is a feasible genetic event given that most SPATE genes are encoded in plasmids or flanked by IS-like elements within unstable pathogenicity islands in the

chromosome. Recently, it was found that the deadly German outbreak EAEC C227-11 strain, which caused at least 50 deaths in Europe in 2011 [71, 72], exhibited the three SPATEs produced by *Shigella*, SigA, SepA, and Pic, rarely reported in EAEC strains, but a clear event of lateral transfer of SPATE genes [72]. Evidence of other genetic events, including recombination, spontaneous point mutations, or deletions, may have also contributed to the vast diversity of SPATEs [99–101].

Functions of SPATES and roles in pathogenesis

Previously, it was believed that there is no specific correlation between phylogenetic groupings and the biological function of SPATEs [8]. This assumption was based on the fact that, despite members of the same class having similar substrates, they do not share identical target peptide cleavage specificities [38, 56]. However, new reports on SPATE substrate-specificity and biological activities suggest that there is a tendency toward functional correlation among the proteins and their clustering in the phylogenetic tree [61, 64, 67, 102, 103] (Fig. 5; Table 1). Regardless of their substrate or cleavage sites, class-1 SPATEs, for example, have a common ability to cause cytopathic effects in cultured cells and display enterotoxin activity [63, 64, 67, 89, 90, 93, 95, 103] (Fig. 6a; Table 1), whereas most studied class-2 SPATEs exhibit a lectin-like activity with predilection to degrade a variety of mucins, including leukocyte surface O-glycoproteins with vital roles in numerous cellular functions, resulting in advantage for mucosal colonization and immune modulation [56, 61, 68, 70, 75, 104] (Fig. 6b; Table 1). The fact that there was no reliable animal model for most of the human pathogens harboring SPATEs, the majority of SPATE functional assays and their role in pathogenesis, with few exceptions, have been carried out in in vitro, in situ, and ex vivo settings.

Class-1 SPATEs, the cytotoxic serine proteases

Although the first SPATE discovered almost two decades ago belongs to the class-2 subfamily, class-1 SPATEs have been more extensively studied in terms of function and biological activities.

All class-1 SPATEs studied so far have displayed cytotoxic effects on cultured cells (Fig. 6a; Table 1), and enterotoxin activity on intestinal tissues (as determined by the fluid accumulation in ligated intestinal loops, and rises in the short-circuit current of the intestinal tissue mounted in Ussing chambers). Well-defined functional studies have been performed for EspP (extracellular serine protease plasmid-encoded), PssA (protease secreted by STEC), Pet

Fig. 6 Effect of Class-1 and Class-2 SPATEs on host cells. **a** Effects ▶
of SPATEs on HEp-2 cell monolayers are shown by oil immersion
light microscopy of Giemsa-stained HEp-2 cells after treatment with
SPATE proteins at 500 nM for 5 h. Rounding of cells is seen with Pet
and Sat (*arrows*). Obvious cytotoxic effects of the less active EspP
and EspC on cells are only seen with higher protein concentrations
and exposure times (see text). Copyright© 2002, American Society
for Microbiology [56]. **b**. Pic produced from *Shigella flexneri* and
pathogenic *E. coli* degrades a broad range of leukocyte glycoproteins.
To visualize degradation of O-linked glycoproteins by Pic, whole
blood leukocytes were treated with Pic and the protease defective
PicS258A for 30 min, stained for DNA (*blue*) and PSGL-1(*red*), and
analyzed by confocal microscopy. Complete degradation of PSGL-1
can be observed on the Pic-treated leukocyte population. Only deg-
radation of PSGL-1 is shown, but the array of Pic targets included
CD43, CD44, CD45, CD93, and CX3CL1; illustrated in (**c**)

(plasmid-encoded toxin), EspC (EPEC secreted protein C),
Sat (secreted autotransporter toxin), and SigA (*Shigella*
IgA-like protease homologue).

EspP and PssA were discovered at the same time by two
different research groups, but represent the same protein
[92, 93]. EspP/PssA is widespread among clinical EHEC
and STEC isolates of the serogroups O113, O157, O26,
O111, and O145 [92, 93, 99, 105], and was the first class-1
SPATE protein shown to have cytotoxic effects on eukary-
otic cells [93]. EspP/PssA-induced cytoskeletal damage
was characterized by the loss of stress fibers, disruption
of the actin cytoskeleton, cell-rounding, cell-detachment,
and opening of cell tight junctions in Vero cells [93].
While authors observed cytotoxicity in Vero cells exposed
to EspP/PssA in the initial study, other authors did not
observe these cytotoxic effects in further epithelial cell
lines (HT-29 and HEp-2), or by other exposure times (30
or 48 h) with 1 µM EspP [106]. Similarly, another group
found cytopathic but not cytotoxic effects in HeLa cells
after prolonged incubation (24 h) or higher concentrations
exposures with EspP [51]. In our laboratory, we found sim-
ilar results; no cytotoxic effect was seen when incubating
HEp-2 and Vero cells with 1 µM of EspP for 5 h (Fig. 6a)
[56]. However, this discrepancy in results is not hard to
reconcile if we observe the variable cytotoxic effect elic-
ited by other class-I SPATEs [56, 64, 65, 67, 94] (discussed
below), which also have distinctive degradation patterns on
α-spectrin—the target being linked to the cytotoxic effects
[56, 64, 65]. Consequently, amino acid variation among
SPATEs, protease doses, time of exposure, mammalian cell
types, and bacterial strains from where SPATEs are purified
[107, 108] may have influenced the strength of protease
activity.

In another study, EspP was found to cleave the human
coagulation factor V, and was proposed to be involved in
the mucosal hemorrhage observed in patients with hemor-
rhagic colitis during EHEC infection [92]. EspP was also
shown to cleave swine pepsin A [92] and human apoli-
poprotein A-I [88], but the biological relevance of these
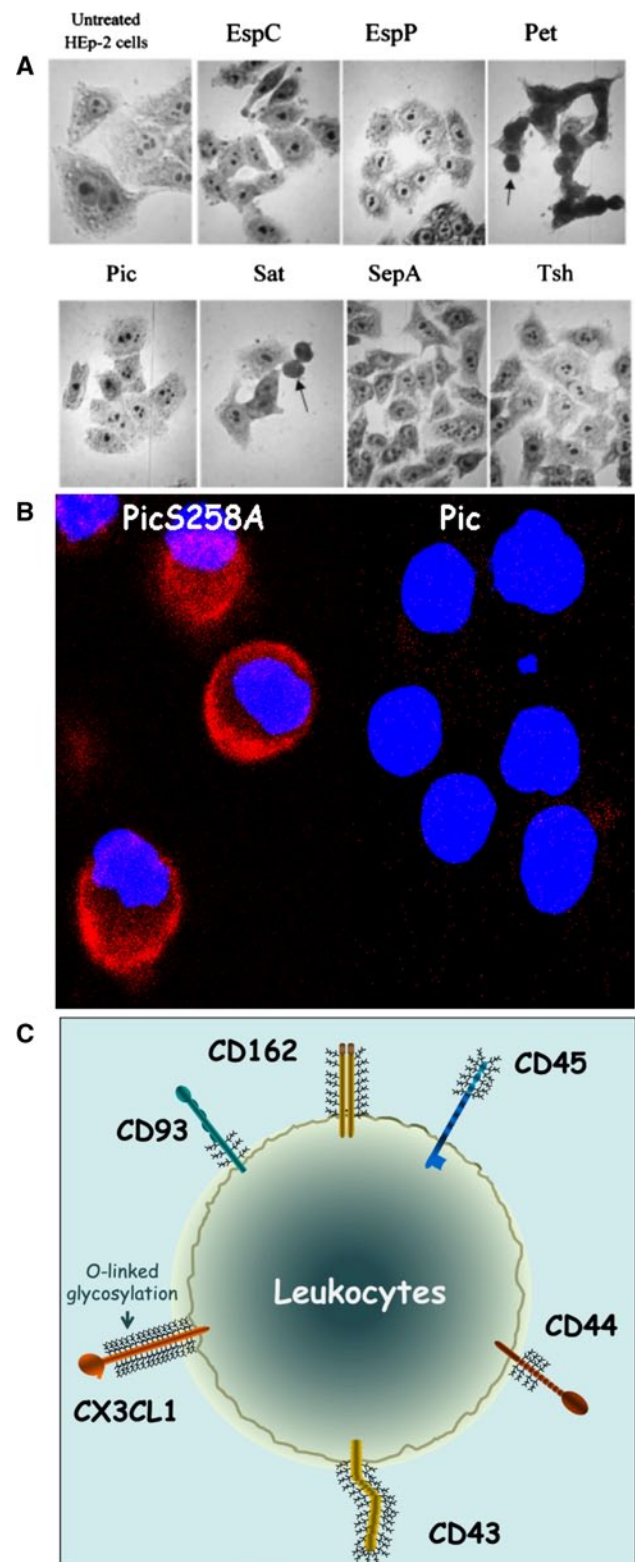
activities was not further investigated. In a subsequent
study, EspP was shown to cleave complement factors C3/
C3b and C5 [109]. C5-depleted serum that was supple-
mented with purified C5 pre-incubated with EspP showed
significantly reduced complement activation in all three

activation pathways (the classical, the alternative, and the mannan-binding lectin pathways). Consequently, it was suggested that complement depletion might protect EspP-secreting EHEC from opsonization, complement-mediated lysis, and inflammatory events [109].

The role of EspP in the pathogenesis of EHEC was demonstrated in cattle, the key reservoir and source of EHEC infections in humans. Signature-tagged mutagenesis in the EHEC O26:H strain, which belongs to serogroups predominantly associated with human infection in developed countries, identified EspP as one of the genes required for intestinal colonization of calves [110]. Moreover, EspP was later shown to contribute in adherence to bovine primary rectal cells and colonization of the bovine intestines by *E. coli* O157:H7. In this study, the EspP mutant was shed in feces of colonized animals at lower numbers compared to the parent strain, and the attenuation was statistically significant and comparable to the inactivation of other important EHEC virulence factors [111, 112]. The participation of EspP in bacterial adhesion was reinforced by a transposon mutagenesis study performed in the EHEC EDL933 strain (O157:H7), which identified EspP as one of the virulence factors directly involved in biofilm formation and adherence to T84 intestinal epithelial cells [111]. Recently, it was shown that EspP is able to polymerize and form "rope-like structures," which correlates with its cytopathic effects on cultured epithelial cells; the structures may also serve as a substratum for bacterial adherence and biofilm formation and may also protect bacteria from antimicrobial compounds [51]. An excellent review covering more detailed information on the function of EspP is now available [106]. Following EspP discovery, Pet, another class-1 SPATE, was found in clinically relevant pathogenic *E. coli* strains [89].

Pet was originally found encoded in the pAA2 virulence plasmid of the prototype EAEC strain 042 and is predominantly distributed in several other clinical EAEC isolates [89, 102]. Pet is perhaps the most comprehensively studied class-1 SPATE to date. Initial studies identified Pet as a 108-kDa heat-labile enterotoxin and cytotoxin by virtue of its ability to increase short-circuit currents (Isc) and decrease electrical resistance of the rat jejunal tissue mounted in Ussing chambers [89], which was often accompanied by mucosal damage, exfoliation of cells, and development of crypt abscesses [63]. Subsequently, purified Pet protein was found to induce cytopathic effects in HEp-2 and HT29 C1 cells, a phenotype dependent on the proteolytic activity mediated by the catalytic serine protease motif in the Pet passenger domain [94]. Subsequent experiments have shown that epithelial cells exposed to purified Pet display phenotypic changes after 2 h of toxin exposure, and that exposure times as short as 10 min are sufficient for inducing cytopathic effects [94]. The cytopathic effects triggered by Pet were similar to the ones seen with EspP intoxication

in the early study [93]. Thus, elongation, rounding, and detachment of the cells from the substratum; cytoskeleton contraction; loss of actin stress fibers, and release of focal contacts are the classical signals of intoxication induced by class-1 SPATEs [89, 94, 102] (Fig. 6a). Furthermore, it was shown that cell intoxication was associated with the ability of Pet to cleave the actin-binding protein α-fodrin (α-spectrin), as demonstrated using in vitro and in vivo settings [66, 113, 114]. Interestingly, these studies showed that Pet cleavage on α-fodrin occurred within the calmodulin-binding domain of fodrin's 11th repetitive unit, which led to cytoskeleton disruption [66]. Pet was also found to cleave additional cytoskeletal elements in HEp-2 and HT29 cells, such as focal adhesion kinase (FAK), a protein involved in focal adhesion complexes. Following treatment with Pet, FAK was tyrosine dephosphorylated, before the redistribution of FAK and spectrin occurred. Moreover, phosphatase inhibition blocked cell retraction, suggesting that tyrosine dephosphorylation is an event that precedes FAK cleavage [115].

Trafficking studies have revealed that Pet was internalized into host cells via clathrin-dependent endocytosis [116]. Once inside the cells, Pet was found to move from the cell surface to endosomes, Golgi apparatus, endoplasmic reticulum, and, ultimately, Pet was delivered back to the cytosol to reside in close contact with its substrate, the actin-binding protein α-foldrin (spectrin) [117]. Importantly, Pet represented the first bacterial toxin found to target α-fodrin and the first SPATE to display enterotoxin activity [89, 114]. Following Pet's discovery, many other class-1 SPATEs were found to cleave α-fodrin and to trigger similar biological effects. Given that there is no suitable animal model which reproduces EAEC infection, the role of Pet in pathogenesis has been inferred from human colonic biopsy specimens maintained in an in vitro organ culture (IVOC) model [63]. Using this model, the EAEC *pet* mutant strain exhibited significantly fewer mucosal abnormalities than the parental EAEC 042 strain, characterized by exfoliation of enterocytes, dilation of crypt openings, increased cell rounding, development of prominent inter-crypt crevices, and absence of apical mucus plugs. Moreover, the mucosal effects were restored upon trans-complementation with *pet* [63]. Following Pet findings, three more SPATEs, EspC, Sat, and SigA, were identified in relevant pathogenic *E. coli* and *Shigella* strains.

EspC is a 110-kDa protein originally found within a second pathogenicity island on the chromosome of the prototype EPEC E2348/69, but not in all EPEC isolates [95]. EspC was found located in a DNA region with a substantially lower G/C content than the rest of the EPEC chromosome, along with genes similar to a variety of mobile genetic elements; thus, it was postulated that EspC was more likely acquired by horizontal transfer [95]. Like Pet,

EspC showed enterotoxin activity in the rat jejunal tissue mounted in Ussing chambers [95]. In the same study, antibodies directed against the Pet protein of EAEC cross-reacted with EspC in western blots, and pre-incubation with Pet antiserum inhibited EspC's enterotoxin activity in Ussing chambers, suggesting parallel substrates and biological activities to the Pet toxin. Indeed, in a later study, EspC was also found to elicit cytotoxic effects in cultured epithelial cells, characterized by cell contraction and detachment from the substratum due to disruption of the actin cytoskeleton [65]. Nevertheless, efficient intoxication with EspC was reached only when higher concentrations and prolonged exposure of EspC were used [65]. In such studies, purified EspC was internalized by a non-specific pinocytic mechanism, without an apparent membrane receptor since EspC did not bind to the plasma membrane at 4 °C [65]. Subsequently, it was found that efficient internalization and intoxication with EspC required EPEC contact with the host cell, and a functional EPEC type III secretion machinery [118]. EspC was efficiently delivered to the host cell cytosol when cultured cells were exposed to the EPEC type III secretion system (T3SS), which accelerated EspC delivery to the host cytosol within 1 h of infection. Moreover, EPEC mutants with defective T3SS were unable to translocate either endogenous or exogenous EspC into the epithelial cells [118, 119]. Despite these observations, the precise mechanism of T3SS delivery of EspC to the host cells remains unknown.

EspC also cleaves α-fodrin, an effect dependent on a functional serine protease motif. Unlike Pet, EspC cleaves at fodrin's 11th and 9th repetitive units outside of the calmodulin-binding domain, which may explain the different cytopathic and enterotoxic effects seen with EspC intoxication, such as the absence of the classical redistribution of foldrin proteolytic fragments into membrane blebs seen in Pet intoxication [65]. EspC has been shown to cleave many other substrates including pepsin, hemoglobin, and coagulation factor V [56, 120], and, since EspC is a hemin-binding protein, it has been hypothesized that hemoglobin proteolysis by EspC may contribute to the utilization of heme and hemoglobin iron for bacterial growth [120], a virulence trait that could be enhanced by the inactivation of coagulation factor V. However, the biological relevance of these observations in vivo remains to be established.

EPEC infection induces characteristic "attaching and effacing" (A/E) lesions on the intestinal mucosa, which are defined by the effacement of microvilli and the formation of an actin pedestal underneath the bacterium [7]. Originally, EspC was found to be associated with HeLa cells in vitro without any involvement in the formation of the classical A/E lesions [121]. Like Pet and EspP, EspC may have an accessory role in the initial phase of EPEC infection, where it would promote intestinal colonization and cytotoxic effects, and perhaps enhance diarrhea, in a disease characterized by profuse watery, and sometimes bloody, diarrhea.

A suitable animal model for EPEC infection does not yet exist, instead, *C. rodentium*, a natural mouse pathogen that causes (A/E) lesions similar to those observed with EPEC and EHEC, is used to model pathogenesis [122]. The in vivo role of EspC in pathogenesis remains inconclusive; however, with the discovery of CrC1sp, an EspC-homologue in *C. rodentium*, it is now possible to address its role in bacterial pathogenesis (Fig. 5).

SigA is a 103-kDa protein originally found in the *she* pathogenicity island (PAI) of *Shigella flexneri* 2a [90, 123]. Interestingly, SigA was located along with two other autotransporter proteins within the same *she* PAI: Pic and an Ag43-like autotransporter [90, 123]. SigA was found to have significant cytopathic effects on HEp-2 cells and contributed to intestinal fluid accumulation in a rabbit ileal loop model of infection [90]. Likewise, the other class-1 SPATEs, SigA, was able to degrade recombinant human α-fodrin in vitro, at the same cleavage site as does Pet, suggesting that the cytotoxic and enterotoxic effects mediated by SigA are likely associated with the degradation of epithelial cell α-fodrin [64]. In contrast to EspC, which did not produce the classical redistribution of α-foldrin seen following Pet intoxication, SigA did cause fodrin redistribution in the host cell; however, when compared to Pet, purified SigA toxin induced a less pronounced level of toxicity, which may result from differences in their binding affinities to α-fodrin [64].

Shigella spp. are the etiological agents of bacillary dysentery. In 1999, it was estimated that *Shigella* causes the death of more than 1.1 million people annually [124, 125]. After ingestion, *Shigella* invade the colonic epithelium and then spread from cell to cell, resulting in cell destruction, inflammation, and watery diarrhea, that often progresses to bloody, mucoid diarrhea [126]. A variety of virulence determinants play important roles in the pathogenesis of shigellosis. SigA enterotoxin may contribute to the watery diarrhea, cell destruction, and perhaps the inflammatory process. As observed with Pet, sera from a convalescent individual infected with *Shigella flexneri 2a* exhibited high anti-SigA titers, suggesting that SigA is produced in vivo [64]. Moreover, SigA was found to be over expressed by *Shigella* at 37 °C, the physiologic temperature [90]. The precise role of SigA in vivo requires further investigation in suitable animal models.

More recently, the *sigA* gene was also found among relevant clinical EAEC isolates, including the EAEC outbreak C227-11 strain [71, 72]; SepA and Pic were also found in this strain. Perhaps the number and combination of SPATEs in the outbreak strain may have contributed to its heightened virulence.

Sat, a 107-kDa protein, was first identified in the UPEC CFT073 strain, but later also found in clinical isolates of *Shigella*, EAEC, and DEAC species [71, 91, 127]. Although the *sat* gene has been found in *Shigella* species [127], *Shigella flexneri* 2a harbored a truncated form of *sat* with the passenger and translocator domains distantly separated within the *Shigella* chromosome, suggesting that a chromosomal rearrangement event occurred in the *sat* locus [128]. However, *Shigella dysenteriae*, the more aggressive *Shigella* species, has an intact *sat* gene, identical to the *sat* from UPEC [129]. Like EspC and SigA, Sat is also located in a pathogenicity island with a significantly lower G/C content than the rest of the UPEC chromosome, suggestive of a lateral transfer event [91]. Sat from both UPEC and DEAC strains has been studied in detail in cell lines and animal models. In the initial study, Sat protein derived from *E. coli* CFT073 exhibited serine protease activity, as well as cytopathic effects, in VERO, HK-2, and HEp-2 cells [91] (Fig. 6a). However, the *sat* mutant in a murine ascending urinary tract infection model (UTI) did not reveal differences in colonization compared to its parent, as assessed by the number of bacteria in the urine, bladder, and kidney [91]. Nevertheless, Sat induced a strong antibody response in mice infected with the wild-type strain [91]. In a more detailed study using relevant cell lines to the urinary tract, such as those derived from the bladder and kidney epithelium, Sat elicited elongation of cells with apparent impairment of cellular junctions upon incubation with the kidney cells [130]. Surprisingly, in contrast to other class-1 SPATEs, incubation with Sat triggered vacuolation of the cytoplasm of human bladder (CRL-1749) and kidney (CRL-1573) cell lines. Although no reduction of CFU in urine, bladder, or kidney tissue was seen following transurethral infection of CBA mice with a *sat* mutant, when compared to the wild-type UPEC CFT073, significant histological changes, including dissolution of the glomerular membrane and vacuolation of proximal tubule cells were seen within the kidneys of mice infected with wild-type UPEC CFT073, but not in the kidney sections of mice infected with the Sat mutant [130], suggesting possible contribution of Sat to the damage of kidney epithelium during upper urinary tract infection by UPEC. Further studies revealed Sat to be internalized by host cells and localized in the cell cytoskeleton, where Sat degraded the typical class-1 SPATE target α-fodrin and several other targets, including the leukocyte function-associated molecule-1 (LFA-1) and nuclear proteins [56, 67]. As with many other class-1 SPATEs, Sat was found to also degrade the human coagulation factor V; however, it did not cleave pepsin [56]. Moreover, Sat induced rearrangements of tight junction (TJ)-associated proteins, ZO-1, ZO-3, and occludin, and increased the paracellular permeability of enterocyte-like Caco-2/TC7 cell monolayers [131]. Nevertheless, whether or not the TJ proteins were cleaved by Sat was not demonstrated. Recently, the vacuolation of the cell cytoplasm was found related to the ability of Sat to induce autophagy before triggering cytoskeleton disruption and cell detachment [132]. Autophagy is a tightly regulated process involving the degradation of a cell's own components through the lysosomal machinery [133]. This process plays a normal part in cell growth, development, and homeostasis, but when intentionally triggered, such as the result of Sat activity, it could represent a bacterial virulence strategy to harm host cells. It would be interesting to establish if this mechanism of cell death triggered by Sat is actually the mechanism of cytotoxicity induced by class-1 SPATES. Given that the pathogenicity of UPEC implies the exfoliation of infected bladder epithelial cells, it has been hypothesized that Sat, by inducing autophagic cell detachment, may be involved in cell exfoliation during UTI caused by UPEC [132]. Sat has also shown enterotoxin activity in rabbit ileum tissues in Ussing chamber assays, as well as pronounced fluid accumulation and villous necrosis in rabbit ileum loops [103]. The biological relevance of this observation could be envisioned from the diffusely adhering *E. coli* (DAEC) strains, which are responsible for urinary tract and intestinal infections. Sat was found not only in the urine of patients with UTI but it was also highly prevalent in clinical collections of Afa/Dr DAEC strains recovered from the stools of children with active diarrhea [131].

Other class-1 SPATEs

Most of the hypothetical class-1 SPATEs identified from a search in GenBank were allelic variants of the EspC cytotoxin (Fig. 5; Table 1). None of them have yet been studied in terms of expression, secretion profiles, and the effect on eukaryotic cells. However, their similarity to EspC leads us to believe that they may have substrate predilection, protease activity, and roles in pathogenesis similar to those of other class-1 cytotoxins. As noted above, EspC may be delivered by EPEC T3SS into host cells [118]. The SPATEs termed here as RE22-C1sp and Cr-C1sp are present in animal pathogens (REPEC and *C. rodentium*) which encode the T3SS machinery, and which reproduce crucial aspects of intestinal infections seen with EPEC. Therefore, these pathogens are suitable models to study the SPATE toxins in the context of a natural intestinal infection.

Class-2 SPATES, the lectin-like immunomodulatory serine proteases

Little was known regarding the biological substrates, potential mechanism of action, and possible role in pathogenesis for most of the class-2 SPATEs. Nevertheless,

recent findings revealed new unsuspected biologically significant substrates and potential virulence roles for these proteins. However, more studies in new animal models are required to re-examine the actual roles of these proteases in their respective pathogens. Most class-2 SPATEs studied to date have displayed proteolytic activity on mucins from different sources, more likely by virtue of their ability to adhere glycoproteins, which may also reflect their ability to agglutinate red cells [69, 134, 135]. Initially, SPATEs with mucinolytic activity were associated with intestinal colonization, thought due to metabolic competitive advantage conferred by the mucus cleavage [68]. We recently showed that prototype class-2 SPATEs bound and cleaved a variety of leukocyte surface glycoproteins with diverse roles in numerous cellular and immune functions, and which were substituted with carbohydrates structurally similar to those found on human mucin glycoproteins [61]. More interestingly, the serine protease activity enhanced leukocyte impairment, but SPATE binding to cells was sufficient to significantly cause a number of effects in leukocytes. Consequently, a new role of class-2 SPATEs suggests its involvement in immune evasion [61].

Pic (protease involved in intestinal colonization), Tsh (temperature-sensitive hemagglutinin), and Hbp (hemoglobin binding protein) are the class-2 SPATEs from which more functional data are available [69, 79, 135]. In addition, some studies have been carried out for PicU (Pic from uropathogenic *E. coli*) [75, 80], Vat (vacuolating autotransporter toxin) [80], EatA (ETEC autotransporter A) [87], EpeA (EHEC plasmid-encoded autotransporter) [70], EspI (*E. coli* secreted protease island-encoded) [88], SepA (*Shigella* extracellular protein A) [136], AdcA (adhesin involved in diffuse *Citrobacter* adhesion)[57], and RpeA (REPEC plasmid-encoded autotransporter) [39].

Tsh/Hbp

Tsh was the first SPATE discovered by virtue of its hemagglutinin activity nearly two decades ago, from the avian pathogenic *E. coli* (APEC) strain ×7122— a pathogen which causes respiratory tract lesions and septicemia in poultry [135]. When expressed in *E. coli* K12, Tsh was also found to confer a hemagglutination phenotype. The initial studies found that rather than the secreted form of Tsh, low levels of unprocessed, cell-associated Tsh, conferred hemagglutination [135]. However, soon after, the secreted form of Tsh was also found to agglutinate red cells and bind to hemoglobin and extracellular matrix proteins, such as fibronectin and collagen IV, and the agglutination phenotype was independent of its protease activity given that replacement of the catalytic serine residue with alanine did not abolish binding to red cells [78, 134]. Tsh provided

the first evidence that a "lectin-like" activity was present in class-2 SPATEs.

Hbp, initially identified in the *E. coli* strain (EB1) [79] isolated from a patient presenting a wound infection, differs from Tsh in only two amino acids; therefore, these proteins represent the same SPATE in different pathogens, and in some instances they are referred here as Tsh/Hbp. Initial studies identified Hbp as a hemoglobin binding protein from where its name was derived [79]. Hbp was able to interact with hemoglobin, degrade it, and subsequently bound the released heme. This fact led to the hypothesis of the involvement of Hbp in heme acquisition by the *E. coli* EB1 pathogen [79]. In agreement with these findings, Hbp was indeed found to contribute as an iron source generator in an infection synergy mouse model between *E. coli* and *Bacteroides fragilis*, where mice immunized with Hbp were protected against mixed infections and did not develop abscess lesions. Likewise, *E. coli* which did not express Hbp was not able to assist abscess formation [137]. Moreover, purified Hbp was able to deliver heme to *B. fragilis* strain BE1 in growth-promoting experiments in vitro [137]. On the other hand, experimental inoculation of chicken with APEC and an isogenic *tsh* mutant demonstrated the contribution of Tsh to the development of lesions within the air sacs of birds, but no clear difference was seen in subsequent generalized infections, such as perihepatitis, pericarditis, and septicemia, suggesting a possible role of Tsh in early stages of infection [138]. Interestingly, Tsh/Hbp is highly prevalent in extraintestinal *E. coli*, mostly in those capable of causing disease in birds [139]. Tsh was found in more than 50 % APEC isolates and particularly more frequent in high-lethality isolates compared to low-lethality isolates [138]. Likewise, analysis of the prevalence of autotransporter genes among 295 APEC strains, previously classified for virulence based on lethality for 1-day-old chickens, demonstrated that *tsh* and *vat* sequences were strongly associated with the high-virulence lethality class [139]. Thus, it appears that Tsh/Hbp contributes to at least two different infectious diseases: intra-abdominal infections (IAI) in humans and respiratory tract infections in poultry.

As mentioned before, Tsh/Hbp has the ability to degrade hemoglobin as an iron source, to cleave mucin, and hemagglutinate red cells. Whether these are the virulence attributes exhibited by Tsh/Hbp in vivo during the natural course of infection are unknown.

We recently reported that, like the Pic serine protease, Tsh/Hbp efficiently bound and cleaved a broad range of O-linked glycoproteins present on the surface of neutrophils and lymphocytes, and that cleavage of these glycans triggered adverse effects on leukocytes [61]. Nevertheless, more refined studies are required to further characterize whether these effects occur in vivo. Interestingly, the

closest homologue of Tsh is Vat, and, like Tsh, Vat is broadly distributed among strains capable of causing septicemia in birds and humans. Hence, it is tempting to speculate that the main role of class-2 SPATEs in extra-intestinal *E. coli* infections is perhaps during the later stages of disease; for example, once bacteria have reached the bloodstream, where SPATEs may impair immune functions, and consequently, pathogens will have an extra advantage to survive and cause meningitis. As early proposed for class-2 SPATEs, Tsh/Hbp was targeting extracellular substrates. The Tsh/Hbp predilection to O-glycoproteins most likely was due to its agglutinin activity. Specific carbohydrates present on red cells are also present on leukocytes, then the "lectin" region of Tsh/Hbp may have interacted with carbohydrates in leukocyte glycans, making them vulnerable to digestion. Further studies are required to fully characterize the role of Tsh/Hbp in immune evasion.

Vat

Vat, a 111.8-kDa protein, is encoded in the VAT pathogenicity island (VAT-PI) of the APEC strain Ec222 [80], a strain causing a variety of syndromes including respiratory disease, swollen head syndrome, cellulitis, and septicemia [140]. Subsequently, Vat was found broadly distributed among human ExPEC including UPEC and invasive *E. coli* causing neonatal septicemia and meningitis [76]. The *vat* gene in ExPEC strains have been arbitrarily annotated as *hbp* or *tsh*, yet the SPATE sequence in ExpEC has 96.5/97 % identity/homology to Vat and 69.6/79 % identity/homology to Tsh/Hbp, and, consequently, we have collectively termed them here as Vat-ExEc SPATEs (Vat from ExEc) (Fig. 5).

Surprisingly, Vat exhibited vacuolating activity in cultured cells, which resembled that observed in intoxication with Sat [80, 130]. However, the Vat protein did not cleave casein (the universal substrate for most proteases), while other known class-2 SPATE substrates were not tested. Moreover, the *vat* deletion mutant of Ec222 showed no virulence in respiratory and cellulitis infection models of disease in broiler chickens [80]. More peculiar is the fact that Vat exhibited two amino acid changes (ATSGSPL) in the catalytic GDSGSPL signature of SPATES, which may have inactivated its proteolytic activity. Inactivation of SPATEs by naturally occurring mutations in the catalytic site has been observed before [99]. Moreover, analysis of the Vat homologue in all sequenced Ex-PEC strains has shown an intact GDSGSPL motive (NCBI Accession nos. NP_752330.1 and ADE90928.1). Additional studies are necessary to elucidate the foundation of vacuolating activity, and whether Vat was internalized in host cells, and, if so, whether a non-proteolytic Vat is able to trigger

autophagy similar to Sat, the other SPATE known to vacuolize cells [130]. In UPEC, Vat-ExEc along with PicU were expressed by bacteria isolated from urine of transurethrally-infected mice, where Vat-ExEc protease was identified in 61 % of pyelonephritis and 65 % of cystitis isolates [76]. However, the virulence role of Vat-ExEc in animal models was not investigated.

Pic

Pic, a 109.8-kDa extracellular protein, was originally identified in *Shigella flexneri* 2a and EAEC species [69]. Subsequently, its close homologue PicU was found in UPEC strains [75]. More recently, we have identified a hypothetical Pic-like SPATE, here termed Cr-C1sp in *C. rodentium* (Fig. 5). Originally, Pic was found to possess mucinolytic activity on murine and bovine mucin, which was also seen with PicU [75]. Furthermore, Pic showed hemagglutinin activity in erythrocytes from a variety of species, suggesting an adhesive property, previously shown by the Tsh SPATE [69, 134, 135]. Pic also cleaved the coagulation Factor V, and protected *E. coli DH5α* from complement killing, more likely as a result of the impairment of the classical complement pathway [69]. Unlike class-1 SPATEs, Pic did not cleave pepsin or α-spectrin, and did not display cytotoxic effects on cultured cells [56, 69]. It was then proposed that the main role of Pic in pathogenesis may lie in its mucinolytic activity since the pathogenesis of both *E. coli* and *S. flexneri* infections requires contact with the mucosal cell surfaces, where the mucus layer on the mucosal surface acts as protective barrier against enteric infections. By expressing Pic or Pic-like SPATEs, enteric pathogens may have the ability to penetrate this gel-like layer. Indeed, from studies in a streptomycin-treated mice, an animal model suitable for intestinal colonization studies but not for EAEC disease progression since these animals do not develop diarrhea, Pic promoted EAEC colonization of mouse in a competitive infection assay employing wild-type EAEC042 and its isogenic Pic mutant. Substantially low numbers of the EAEC042pic mutant were found in the lumen, mucus layer, and tissues. Moreover, the wild-type strain exhibited a growth advantage over EAEC042PicS258A (a protease deficient Pic) in a culture of cecal mucus and in cecal contents in vitro, suggesting a metabolic role for the Pic mucinase in EAEC colonization [68]. Given that less colonization of mouse was seen with the Pic mutant than with the wild-type EAEC secreting Pic in the competitive assay, it was possible that the hemagglutinin activity of Pic promoted adhesion to the intestinal tissue or the mucus layer. Adhesion mediated by class-2 SPATEs has been suggested since the initial discovery of SPATEs, where perhaps residual unprocessed Tsh on bacterial surfaces promotes binding of the bacterium to red

cells [135]. This is also seen with AdcA, a newly identified class-2 SPATE from *C. rodentium*, which mediated adhesion to HEp-2 cells [57]. Indirect adhesion of bacteria through "rope-like" structures by class-I SPATEs such as EspP and EspC has also been reported [51]. However, whether class-2 SPATEs mediate adhesion through direct binding to host cells or do they indirectly mediate adhesion by the formation of "rope-like structures has not been investigated.

Interestingly, in rat ileal loops, Pic was found to trigger hypersecretion of mucus, and increase the number of mucus-containing goblet cells in the tissue, which suggested possible participation of Pic in the pathologic features of diarrhea induced by EAEC and the mucoid diarrhea seen in *Shigella* infections [141].

We recently reported a potential new virulence role for Pic and perhaps for all other class-2 SPATEs with the ability to agglutinate red cells (lectin activity). While studying the mechanism of mucin degradation, we discovered that Pic was able to bind and cleave a remarkable array of leukocyte molecules present on the surface of nearly all lineages of hematopoietic cells, including macrophages, neutrophils, and lymphocytes. These targets included CD93, CD43, CD44, CD45, CD162, and the chemokine fractalkine, CXCL1 [61] (Fig. 6b, c). Our data showed that Pic proteolytic activity was dependent on the carbohydrate nature of leukocyte glycans since neuraminidase-treated glycans were not susceptible to Pic protease activity. Nevertheless, removal of N-glycosylation from those targets did not protect from Pic degradation. Moreover, we observed that the key carbohydrate sLeX, found overexpressed in carcinogenic and inflammatory processes, inhibited Pic protease activity, presumably by blocking its lectin domain, which is not yet undefined.

Neuraminidase is known to modify O-glycoproteins by removing terminal sialic acid moieties, and sialic acid is one of the components of the sLeX tetrasaccharide, often present in these glycans. Previously, it was shown that pretreatment of Pic with monosaccharides similar to those found in mucin partially inhibited binding of Pic to mucin [104]. Therefore, it is possible that Pic may also recognize other oligosaccharides present on leukocyte glycans. Additionally, we found that Pic cleaved precisely before Ser and Thr residues—the amino acids linked to carbohydrate branches present in O-glycoproteins.

We also recognized other potential Pic targets, all members of the O-glycoprotein family, including CD34, CD68, CD99, CD164, and the recently discovered and growing family of Tim (T cell, immunoglobulin and mucin domain) proteins, a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity [142]. Nevertheless, they have not been experimentally tested.

Our data also showed that, upon binding to leukocytes, Pic induced activation of neutrophils, but impaired their chemotaxis and transmigration functions. This impairment was in part due to the strong binding of leukocytes to the cell monolayer as a result of the hydrophobic coat loss assembled by O-glycoproteins. In addition, leukocytes also responded weakly to the chemokine IL-8 or fMLP (*N*-formyl-methionyl-leucyl-phenylalanine) as demonstrated in migration experiments without cell monolayers, suggesting impairment of chemotaxis. We observed activation of PMNs following Pic treatment, although binding of Pic to PMN was sufficient to observe significant cell activation, most likely by triggering signal transduction upon binding to glycoproteins. The same phenotype was seen in activated lymphocytes which underwent apoptosis after treatment with Pic or PicS258A (a protease defective Pic). As mentioned before, we also observed the same array of substrates for Tsh/Hbp. In view of the fact that other class-2 SPATEs exhibit mucinolytic and/or agglutinin activity, it is predictable that the majority of class-2 SPATEs may behave in the same manner. Remarkably, O-glycoproteins targeted by Pic are involved in numerous cellular and immune functions. O-glycans support leukocyte recruitment in both innate and adaptive arms including cell migration, cell adhesion, cell–cell and cell–matrix interactions, signal transduction, modulation of cytokine production, phagocytosis, and chemotaxis [143–146].

The precise mechanism of cell activation and apoptosis by Pic is not yet completely understood; however, a possible mechanism of action of Pic on PSGL-1 can be envisioned. PSGL-1 is fundamental in leukocyte trafficking, and the rolling interaction between leukocyte PSGL-1 and endothelial selectins requires branched O-glycan extensions on specific PSGL-1 amino acid residues, such as the sLeX, the same tetrasaccharide recognized by Pic. In neutrophils, the glycosyltransferases involved in formation of the O-glycans are constitutively expressed, while in T cells, they are expressed only after appropriate activation. Interestingly, we observed that Pic triggered the oxidative burst in resting PMNs, but apoptosis was triggered only in activated lymphocytes [61]. It is known that lymphocytes overexpress O-glycosylated PSGL-1 in the activated state, and cross-linking of PSGL-1 with specific antibodies can trigger apoptosis [147]. It is therefore possible that, beyond Pic protease activity, cross-linking of PSGL-1 by Pic may have taken place. Nevertheless, many other Pic targets modified with sLeX have also been implicated in cell activation and apoptosis [148–150].

Targeting of leukocyte O-glycans as a mechanism of virulence exploited by extraintestinal pathogens is easily foreseen; however, the advantage of targeting these glycans by the intestinal pathogens would be more complex to comprehend, particularly because Pic exhibited a dual

activity: it was able to activate PMN upon contact, which may aggravate the inflammatory process, but it also causes apoptosis in activated cells and inhibits leukocyte transmigration, which may lessen the inflammatory process [61]. Indeed, we observed that *S. flexneri* 2a was more inflammatory in the Guinea pig conjunctivitis model than the *Shigella pic* mutant, suggesting an anti-inflammatory role for Pic [61]. Furthermore, recent studies have shown that adhesion molecules, such as PSGL1, CD34, and fractalkine, are important in the intestinal innate immune response against enteric infections [151–153]. Infected mice lacking PSGL-1 or P-selectin with *C. rodentium*, a mouse pathogen which reproduces the disease seen with the human pathogens EPEC and EHEC [122], showed a more pronounced morbidity associated with higher bacterial load, elevated IL-12 p70, TNF-alpha, IFN-gamma, MCP-1 and IL-6 production, and more severe inflammation than the wild-type mouse [153, 154]. Likewise, infected mice lacking core 2 (O-glycosylation), PSGL-1, or P-selectin with *Salmonella* showed a more pronounced morbidity and a significantly higher mortality rate associated with higher bacterial load and proinflammatory cytokine production than the wild-type control mouse [129]. In another study, CD34, a highly glycosylated sialomucin promoted leukocyte transmigration in the gut upon infection with *Salmonella*, where increased numbers of CD34+ cells were detected in the submucosa, vascular endothelium, and lamina propria. In contrast, CD34(−/−) mice showed a delayed pathology, a defect in inflammatory cell migration into the intestinal tissue, and enhanced survival [152]. Similar results were seen in fractalkine (CX3C) and in the chemokine receptor (CX3CR1) knockout mice, which displayed markedly increased translocation of commensal bacteria and *Shigella* [155].

Analogous effects may take place in the context of a natural infection with class-2 SPATE producers, which may in theory cleave not one, but an array of leukocyte O-glycoproteins, as shown in vitro. Nevertheless, more rigorous studies are necessary to further characterize the role of Pic and other class-2 SPATEs in pathogenesis.

SepA

SepA is 110-kDa plasmid-encoded protein originally found in *Shigella flexneri* species [136], but later also found broadly distributed in EAEC clinical isolates [71, 156]. A hypothetical SPATE with 89.3 % homology to SepA, here termed as ERN587-sp, is found in the EPEC strain RN587/1, a recent clinical isolate from Brazil (Fig. 5).

SepA was found to hydrolyze several synthetic peptides similar to those recognized by cathepsin G, a serine protease produced by polymorphonuclear leukocytes; however, it did not cleave cathepsine G natural substrates,

such as fibronectin, collagen, or angiotensin [136]. Furthermore, SepA does not seems to cleave any class-1 or class-2 substrates, including mucin, leukocyte O-glycoproteins, α-spectrin, factor V, or pepsin [56, 61]. Neither does it exhibit a cytopathic effect in HEp2 cells, nor is it involved in bacterial invasion, plaque formation, or *Shigella* cell–cell dissemination [56, 157]. However, *sepA* mutant was attenuated in the rabbit model of ligated ileal loop, characterized by reduced fluid accumulation, reduced mucosal atrophy, and decreased tissue inflammation compared with the wild-type strain, suggesting an enterotoxin activity [157]. Indeed, SepA was most strongly associated with diarrhea among the 121 EAEC strains isolated in 2008 as part of a case–control study in Mali of moderate to severe acute diarrhea among children of 0–59 months of age [158]. Recently, the Shigella-*sepA* mutant showed a significant decrease in surface epithelium desquamation and a significantly higher epithelial height in human colonic explants when compared with the wild-type strain, but no differences in terms of epithelial desquamation or epithelial height were seen when non-invasive *Shigella* expressing SepA was used, suggesting that SepA was necessary, but not sufficient, to induce these alterations [159]. Thus, the authors proposed that SepA may play a role in the generation of mucosal damage in an early event of Shigellosis.

Although SepA is clustered within class-2 SPATEs and has roughly 67 % homology to Pic or Tsh/Hbp, which are SPATEs with same substrate predilection, SepA does not seem to share the same substrates, and even more, its targets are not known. The SepA cluster seems to move away from its class-2 peers (Fig. 5).

EatA

EatA is a 110-kDa plasmid-encoded protein and is highly prevalent in clinical ETEC isolates [87]. ETEC strains possess several virulence factors including fimbria adhesins which mediate adhesion to small bowel enterocytes, and induce watery diarrhea by secretion of heat-labile (LT) and/or heat-stable (ST) enterotoxins [7]. EatA shares 80 % homology with SepA, and, like SepA, EatA cleaved synthetic peptides that have previously been defined as substrates for cathepsin G, but does not cleave the actual cathepsin G natural substrates [87]. The initial study showed that ETEC*eatA* mutant retarded fluid accumulation in the rabbit ileal loop model similar to SepA, suggesting that this autotransporter contributes to the virulence of ETEC [87].

Recently, EatA was found to degrade the adhesin glycoprotein EtpA from the same ETEC strain, resulting in modulation of bacterial adhesion and accelerated delivery of the heat-labile toxin, a principal ETEC virulence determinant [160]. Based on this fact, it is tempting to speculate that perhaps members of the SepA cluster have evolved to

acquire predilection to bacterial glycoproteins instead of the host glycoproteins to modulate their virulence. Whether SepA exhibits the same EatA activity on *Shigella* adhesins, or whether SepA would trans-complement the ETEC*eatA* mutant, is not known. SepA may exhibit the same functional activity; after all, the *Shigella* species elaborate multiple enterotoxins including the lethal Shiga toxin.

EpeA

EpeA is a 111.7-kDa protein originally identified in the pO113 region of the large hemolysin plasmid from the LEE-negative EHEC O113:H21 (EH41) [70]. Like other class-2 SPATEs, EpeA exhibited serine protease and mucinase activity, suggesting predilection to O-glycoproteins. Furthermore, EpeA did not have cytopathic effects on the epithelial cells [70]. EpeA has high homology to EspI, a 110.4-kDa protein and the second SPATE identified in a LEE-negative EHEC strain [88], and homology to a hypothetical SPATE, here termed EH250-sp, found in the STEC H250 strain (Fig. 5). EspI exhibited serine protease activity on pepsin A and human apolipoprotein A-I, but mucinase activity or cytotoxicity in cultured cells was not tested. Nevertheless, it has been hypothesized that these SPATEs may contribute to the pathogenesis of the LEE-negative subset of EHEC strains [88].

Class-2 SPATEs lacking the domain 2

The presence of domain-2 in the passenger domain of SPATEs was previously considered a classifying feature of class-2 SPATEs, first identified in the solved Hbp structure [43]. However, among class-2 SPATEs, we identified a subset of hypothetical allelic SPATE variants, which do not exhibit this domain, though they share high homologies to class-2 SPATEs (Figs. 3, 4, 5). These hypothetical SPATEs are mainly distributed in animal pathogens including the rabbit pathogenic *E. coli*, swine pathogenic *E. coli*, and *C. rodentium*. The two members of this SPATE subset functionally tested to date are RpeA (REPEC plasmid-encoded autotransporter) from REPEC and AdcA (adhesin involved in diffuse *Citrobacter* adhesion) from *C. rodentium* [39, 57].

RpeA is a SPATE lacking the cleavage site (NN) between the passenger and translocator domains and therefore the only non-secreted SPATE so far recognized. RpeA was identified by a signature-tagged mutagenesis approach from isolates defective in intestinal colonization [39]. The *rpeA* mutant took longer to colonize in rabbits, was cleared more quickly than wild-type REPEC 83/39, and did not achieve the high bacterial numbers associated with REPEC

83/39 infection, therefore implying a role for RpeA in intestinal colonization. The *rpeA* gene was found in several REPEC isolates but not in other *E. coli* pathotypes. The authors could not confirm expression and subcellular location of the protein but predicted that *rpeA* encodes a 1,228-amino-acid precursor protein with a molecular mass of 135 kDa [39]. RpeA exhibited a significant degree of amino acid sequence identity with several SPATE proteins including Tsh/Hbp and the domain2-less class-2 SPATEs cluster (Figs. 4, 5); however, it also showed homologies with non-SPATE protease autotransporters [39].

Often, newly sequenced SPATE genes have been arbitrarily annotated as AIDA1-like, IgA protease-like, hemoglobin protease-like or EspC-like without considering the closest homology to the functional SPATE domain. AdcA was initially compared with the AIDA1 autotransporter, an adhesin determinant conferring diffuse adherence to many pathogenic *E. coli* strains. In that study, *E. coli* DH5α expressing AdcA exhibited diffuse adherence to HEp2 cells, but no difference was seen when wild-type *C. rodentium* and its isogenic AdcA mutant were compared [57]. Additionally, the AdcA mutant along with the wild-type *C. rodentium* strain were tested in mice, where the number of bacteria in feces showed no significant difference in these strains in their abilities to colonize in mice [57]. Furthermore, co-infection experiments with both strains showed no difference in colonization. Then, it was concluded that *adcA* does not make a significant contribution in colonization of mice by *C. rodentium* [57]. The AdcA passenger domain exhibits the classifying features of SPATEs, such as the conserved motifs in the passenger and translocator domains, including the typical serine protease motif GDSGS, and the presence of the cleavage site NN between the passenger and translocator domains, from which the passenger is incised during secretion (Fig. 1). Furthermore, AdcA also shows the discrete domain-3, proposed here as a new trait of class-2 SPATES (Figs. 3, 4). Since surface exposure of AdcA or its secretion into the culture supernatant was not demonstrated, it is possible that bacterial adhesion to HEp2 cells may be the result of residual unprocessed AdcA on the bacterial surface, such as the cell adhesion phenotype observed with Tsh, particularly when expressed in *E. coli* DH5α, a laboratory strain which often exhibits residual unprocessed SPATEs on the bacterial surface [111, 135],

It would be interesting to determine whether this cluster of SPATEs lacking domain-2 is able to agglutinate red cells, particularly because the domain-2 has been thought to be involved in substrate recognition. Class 2 SPATEs lacking domain-2 are presently being identified in animal pathogens, and functional studies are required to further characterize their role in virulence.

Concluding remarks

Diarrhea caused by enteric infections is a major factor in morbidity and mortality worldwide. More than 3 billion episodes of infectious diarrhea occur each year and 1 million deaths are estimated to be caused by *Shigella* spp. and diarrheagenic *E. coli* pathotypes [161]. Enteric pathogens utilize a variety of sophisticated strategies to colonize the intestinal tract, evade host defenses, multiply, and damage the host. Although these pathogens have acquired a diverse array of virulence factors, which differentiate them into pathotypes, they collectively exhibit a common virulence factor, termed SPATE. These serine proteases have apparently evolved into two classes with predilection to extracellular and intracellular substrates, allowing pathogens to impair the host homeostasis and modulate the immune response as a pathogenicity mechanism to persist in their environments. Class-1 SPATEs, for example, target intracellular substrates, eliciting cytotoxic and endotoxin effects on the host, which may be the result of the ability to trigger the lysosomal machinery for cell autodegradation—autophagy. On the other hand, class-2 SPATEs seem to disrupt mucosal barriers, provide a metabolic advantage, and modulate the immune response by targeting a variety of leukocyte surface glycoproteins, substituted with carbohydrates structurally similar to those found on human mucin glycoproteins, and with vital roles in numerous immune functions. The high prevalence of SPATEs in enterobacteria and the multiple SPATEs often found in a single pathogen suggest that they accomplish important roles in pathogenesis. SPATEs have also been found, with low frequency, in distantly related pathogens and commensal strains, most likely as a consequence of sharing the same intestinal niche with the enteropathogens, which often encode virulence genes in unstable genetic elements. Nevertheless, more sophisticated strategies are required to investigate the in vivo role of SPATE in pathogenesis. Since SPATEs exhibit redundant activities, it will be necessary to study them in individual settings within the natural progression of diseases, a goal that can be accomplished with the discovery of SPATEs in animal pathogens.

References

1. Pallen MJ, Chaudhuri RR, Henderson IR (2003) Genomic analysis of secretion systems. Curr Opin Microbiol 6(5):519–527
2. Henderson IR, Navarro-Garcia F, Nataro JP (1998) The great escape: structure and function of the autotransporter proteins. Trends Microbiol 6(9):370–378
3. Jacob-Dubuisson F, Locht C, Antoine R (2001) Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol Microbiol 40(2):306–313
4. Linke D et al (2006) Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol 14(6):264–270
5. Salacha R et al (2010) The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. Environ Microbiol 12(6):1498–1512
6. Oberhettinger P et al (2012) Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. PLoS ONE 7(10):e47069
7. Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic *Escherichia coli*. Nat Rev Microbiol 2(2):123–140
8. Yen YT et al (2008) Common themes and variations in serine protease autotransporters. Trends Microbiol 16(8):370–379
9. Desvaux M et al (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends Microbiol 17(4):139–145
10. Henderson IR, Cappello R, Nataro JP (2000) Autotransporter proteins, evolution and redefining protein secretion: response. Trends Microbiol 8(12):534–535
11. Henderson IR et al (2000) Renaming protein secretion in the gram-negative bacteria. Trends Microbiol 8(8):352
12. Selkrig J et al (2012) Discovery of an archetypal protein transport system in bacterial outer membranes. Nat Struct Mol Biol 19(5):506–510
13. Jong WS et al (2010) YidC is involved in the biogenesis of the secreted autotransporter hemoglobin protease. J Biol Chem 285(51):39682–39690
14. Ruiz-Perez F et al (2009) Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of Enterobacteriaceae. J Bacteriol 191(21):6571–6583
15. Ruiz-Perez F, Henderson IR, Nataro JP (2010) Interaction of FkpA, a peptidyl-prolyl cis/trans isomerase with EspP autotransporter protein. Gut Microbes 1(5):339–344
16. Purdy GE, Fisher CR, Payne SM (2007) IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. J Bacteriol 189(15):5566–5573
17. Ieva R, Bernstein HD (2009) Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. Proc Natl Acad Sci USA 106(45):19120–19125
18. Sauri A et al (2009) The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. Microbiology 155(Pt 12):3982–3991
19. Ieva R et al (2011) Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain. Proc Natl Acad Sci USA 108(31):E383–E391
20. Jain S, Goldberg MB (2007) Requirement for YaeT in the outer membrane assembly of autotransporter proteins. J Bacteriol 189(14):5393–5398
21. Rossiter AE et al (2011) The essential beta-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. J Bacteriol 193(16):4250–4253
22. Pohlner J et al (1987) Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature 325(6103):458–462
23. Klauser T et al (1993) Characterization of the Neisseria Iga beta-core. The essential unit for outer membrane targeting and extracellular protein secretion. J Mol Biol 234(3):579–593
24. Bernstein HD (2007) Are bacterial 'autotransporters' really transporters? Trends Microbiol 15(10):441–447
25. Jacob-Dubuisson F, Fernandez R, Coutte L (2004) Protein secretion through autotransporter and two-partner pathways. Biochim Biophys Acta 1694(1–3):235–257

26. Dautin N, Bernstein HD (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. Annu Rev Microbiol 61:89–112

27. Leyton DL, Rossiter AE, Henderson IR (2012) From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. Nat Rev Microbiol 10(3):213–225

28. Leo JC, Grin I, Linke D (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos Trans R Soc Lond B 367(1592):1088–1101

29. Sijbrandi R et al (2003) Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular Escherichia coli protein. J Biol Chem 278(7):4654–4659

30. Szabady RL et al (2005) An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. Proc Natl Acad Sci USA 102(1):221–226

31. Junker M, Besingi RN, Clark PL (2009) Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. Mol Microbiol 71(5):1323–1332

32. Peterson JH et al (2010) Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. Proc Natl Acad Sci USA 107(41):17739–17744

33. Renn JP et al (2012) ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. Chem Biol 19(2):287–296

34. Junker M et al (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. Proc Natl Acad Sci USA 103(13):4918–4923

35. Page MJ, Di Cera E (2008) Serine peptidases: classification, structure and function. Cell Mol Life Sci 65(7–8):1220–1236

36. Rawlings ND, Barrett AJ, Bateman A (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40(Database issue):D343–D350

37. Rawlings ND, Barrett AJ, Bateman A (2010) MEROPS: the peptidase database. Nucleic Acids Res 38:D227–233

38. Dautin N (2010) Serine protease autotransporters of Enterobacteriaceae (SPATEs): biogenesis and function. Toxins (Basel) 2(6):1179–1206

39. Leyton DL et al (2007) Contribution of a novel gene, rpeA, encoding a putative autotransporter adhesin to intestinal colonization by rabbit-specific enteropathogenic Escherichia coli. Infect Immun 75(9):4664–4669

40. Dautin N et al (2007) Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. EMBO J 26(7):1942–1952

41. Barnard TJ et al (2012) Molecular basis for the activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter. J Mol Biol 415(1):128–142

42. Tajima N et al (2010) A novel intein-like autoproteolytic mechanism in autotransporter proteins. J Mol Biol 402(4):645–656

43. Otto BR et al (2005) Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli. J Biol Chem 280(17):17339–17345

44. Emsley P et al (1996) Structure of Bordetella pertussis virulence factor P.69 pertactin. Nature 381(6577):90–92

45. Johnson TA et al (2009) Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of Haemophilus influenzae immunoglobulin A1 protease. J Mol Biol 389(3):559–574

46. Meng G et al (2011) Crystal structure of the Haemophilus influenzae Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. EMBO J 30(18):3864–3874

47. Gangwer KA et al (2007) Crystal structure of the Helicobacter pylori vacuolating toxin p55 domain. Proc Natl Acad Sci USA 104(41):16293–16298

48. van den Berg B (2010) Crystal structure of a full-length autotransporter. J Mol Biol 396(3):627–633

49. Celik N et al (2012) A bioinformatic strategy for the detection, classification and analysis of bacterial autotransporters. PLoS ONE 7(8):e43245

50. Oliver DC et al (2003) A conserved region within the Bordetella pertussis autotransporter BrkA is necessary for folding of its passenger domain. Mol Microbiol 47(5):1367–1383

51. Xicohtencatl-Cortes J et al (2010) Bacterial macroscopic rope-like fibers with cytopathic and adhesive properties. J Biol Chem 285(42):32336–32342

52. Klemm P, Vejborg RM, Sherlock O (2006) Self-associating autotransporters, SAATs: functional and structural similarities. Int J Med Microbiol 296(4–5):187–195

53. Nishimura K et al (2010) Role of domains within the autotransporter Hbp/Tsh. Acta Crystallogr D Biol Crystallogr 66(Pt 12):1295–1300

54. Khan S et al (2011) Crystal structure of the passenger domain of the Escherichia coli autotransporter EspP. J Mol Biol 413(5):985–1000

55. Leyton DL et al (2011) Size and conformation limits to secretion of disulfide-bonded loops in autotransporter proteins. J Biol Chem 286(49):42283–42291

56. Dutta PR et al (2002) Functional comparison of serine protease autotransporters of enterobacteriaceae. Infect Immun 70(12):7105–7113

57. Hart E et al (2008) RegA, an AraC-like protein, is a global transcriptional regulator that controls virulence gene expression in Citrobacter rodentium. Infect Immun 76(11):5247–5256

58. Cheney CP et al (1980) Species specificity of in vitro Escherichia coli adherence to host intestinal cell membranes and its correlation with in vivo colonization and infectivity. Infect Immun 28(3):1019–1027

59. Tan C et al (2011) Genome sequence of a porcine extraintestinal pathogenic Escherichia coli strain. J Bacteriol 193(18):5038

60. Arnold K et al (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22(2):195–201

61. Ruiz-Perez F et al (2011) Serine protease autotransporters from Shigella flexneri and pathogenic Escherichia coli target a broad range of leukocyte glycoproteins. Proc Natl Acad Sci USA 108(31):12881–12886

62. Dutta PR, Sui BQ, Nataro JP (2003) Structure-function analysis of the enteroaggregative Escherichia coli plasmid-encoded toxin autotransporter using scanning linker mutagenesis. J Biol Chem 278(41):39912–39920

63. Henderson IR et al (1999) Involvement of the enteroaggregative Escherichia coli plasmid-encoded toxin in causing human intestinal damage. Infect Immun 67(10):5338–5344

64. Al-Hasani K et al (2009) The immunogenic SigA enterotoxin of Shigella flexneri 2a binds to HEp-2 cells and induces fodrin redistribution in intoxicated epithelial cells. PLoS ONE 4(12):e8223

65. Navarro-Garcia F et al (2004) The serine protease motif of EspC from enteropathogenic Escherichia coli produces epithelial damage by a mechanism different from that of Pet toxin from enteroaggregative E. coli. Infect Immun 72(6):3609–3621

66. Canizalez-Roman A, Navarro-Garcia F (2003) Fodrin CaM-binding domain cleavage by Pet from enteroaggregative Escherichia coli leads to actin cytoskeletal disruption. Mol Microbiol 48(4):947–958

67. Maroncle NM et al (2006) Protease activity, secretion, cell entry, cytotoxicity, and cellular targets of secreted autotransporter toxin of uropathogenic Escherichia coli. Infect Immun 74(11):6124–6134

68. Harrington SM et al (2009) The Pic protease of enteroaggregative *Escherichia coli* promotes intestinal colonization and growth in the presence of mucin. Infect Immun 77(6):2465–2473

69. Henderson IR et al (1999) Characterization of pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. Infect Immun 67(11):5587–5596

70. Leyton DL et al (2003) Transfer region of pO113 from enterohemorrhagic *Escherichia coli:* similarity with R64 and identification of a novel plasmid-encoded autotransporter, EpeA. Infect Immun 71(11):6307–6319

71. Boisen N et al (2009) Short report: high prevalence of serine protease autotransporter cytotoxins among strains of enteroaggregative *Escherichia coli*. Am J Trop Med Hyg 80(2):294–301

72. Rasko DA et al (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 365(8):709–717

73. Campanella JJ, Bitincka L, Smalley J (2003) MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinf 4:29

74. Rajakumar K, Sasakawa C, Adler B (1997) Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect Immun 65(11):4606–4614

75. Parham NJ et al (2004) PicU, a second serine protease autotransporter of uropathogenic *Escherichia coli*. FEMS Microbiol Lett 230(1):73–83

76. Heimer SR et al (2004) Autotransporter genes pic and tsh are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. Infect Immun 72(1):593–597

77. Hull RA et al (1999) Virulence properties of *Escherichia coli* 83972, a prototype strain associated with asymptomatic bacteriuria. Infect Immun 67(1):429–432

78. Stathopoulos C, Provence DL, Curtiss R 3rd (1999) Characterization of the avian pathogenic *Escherichia coli* hemagglutinin Tsh, a member of the immunoglobulin A protease-type family of autotransporters. Infect Immun 67(2):772–781

79. Otto BR et al (1998) Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1. J Exp Med 188(6):1091–1103

80. Parreira VR, Gyles CL (2003) A novel pathogenicity island integrated adjacent to the *thrW tRNA* gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect Immun 71(9):5087–5096

81. Parham NJ et al (2005) Distribution of the serine protease autotransporters of the Enterobacteriaceae among extraintestinal clinical isolates of *Escherichia coli*. J Clin Microbiol 43(8):4076–4082

82. Moriel DG et al (2010) Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic *Escherichia coli*. Proc Natl Acad Sci USA 107(20):9072–9077

83. Fookes M et al (2011) *Salmonella* bongori provides insights into the evolution of the Salmonellae. PLoS Pathog 7(8):e1002191

84. Avasthi TS et al (2011) Genome of multidrug-resistant uropathogenic *Escherichia coli* strain NA114 from India. J Bacteriol 193(16):4272–4273

85. Sandt CH, Hill CW (2000) Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of *Escherichia coli*. Infect Immun 68(4):2205–2214

86. Toh H et al (2010) Complete genome sequence of the wild-type commensal *Escherichia coli* strain SE15, belonging to phylogenetic group B2. J Bacteriol 192(4):1165–1166

87. Patel SK et al (2004) Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. Infect Immun 72(3):1786–1794

88. Schmidt H et al (2001) Identification and characterization of a novel genomic island integrated at selC in locus of enterocyte effacement-negative, Shiga toxin-producing *Escherichia coli*. Infect Immun 69(11):6863–6873

89. Eslava C et al (1998) Pet, an autotransporter enterotoxin from enteroaggregative *Escherichia coli*. Infect Immun 66(7):3155–3163

90. Al-Hasani K et al (2000) The *sigA* gene which is borne on the she pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. Infect Immun 68(5):2457–2463

91. Guyer DM et al (2000) Identification of sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. Mol Microbiol 38(1):53–66

92. Brunder W, Schmidt H, Karch H (1997) EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. Mol Microbiol 24(4):767–778

93. Djafari S et al (1997) Characterization of an exported protease from Shiga toxin-producing *Escherichia coli*. Mol Microbiol 25(4):771–784

94. Navarro-Garcia F et al (1999) Cytoskeletal effects induced by pet, the serine protease enterotoxin of enteroaggregative *Escherichia coli*. Infect Immun 67(5):2184–2192

95. Mellies JL et al (2001) espC pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. Infect Immun 69(1):315–324

96. Clarke DJ et al (2011) Complete genome sequence of the Crohn's disease-associated adherent-invasive *Escherichia coli* strain HM605. J Bacteriol 193(17):4540

97. Yen MR et al (2002) Protein-translocating outer membrane porins of Gram-negative bacteria. Biochim Biophys Acta 1562(1–2):6–31

98. Chaudhuri RR et al (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. PLoS ONE 5(1):e8801

99. Brockmeyer J et al (2007) Subtypes of the plasmid-encoded serine protease EspP in Shiga toxin-producing *Escherichia coli*: distribution, secretion, and proteolytic activity. Appl Environ Microbiol 73(20):6351–6359

100. Davis J et al (2001) Evolution of an autotransporter: domain shuffling and lateral transfer from pathogenic Haemophilus to Neisseria. J Bacteriol 183(15):4626–4635

101. Cote JP et al (2012) Identification and mechanism of evolution of new alleles coding for the AIDA-I autotransporter of porcine pathogenic *Escherichia coli*. Appl Environ Microbiol 78(13):4597–4605

102. Navarro-Garcia F, Sonnested M, Teter K (2010) Host-toxin interactions involving EspC and Pet, two serine protease autotransporters of the Enterobacteriaceae. Toxins (Basel) 2(5):1134–1147

103. Taddei CR et al (2005) Secreted autotransporter toxin produced by a diffusely adhering *Escherichia coli* strain causes intestinal damage in animal model assays. FEMS Microbiol Lett 250(2):263–269

104. Gutierrez-Jimenez J, Arciniega I, Navarro-Garcia F (2008) The serine protease motif of Pic mediates a dose-dependent mucolytic activity after binding to sugar constituents of the mucin substrate. Microb Pathog 45(2):115–123

105. Khan AB et al (2009) Serine protease espP subtype alpha, but not beta or gamma, of Shiga toxin-producing *Escherichia coli* is associated with highly pathogenic serogroups. Int J Med Microbiol 299(4):247–254

106. Weiss A, Brockmeyer J (2013) Prevalence, biogenesis, and functionality of the serine protease autotransporter EspP. Toxins (Basel) 5(1):25–48

107. Tapia-Pastrana G et al (2012) VirK is a periplasmic protein required for efficient secretion of plasmid-encoded toxin from enteroaggregative *Escherichia coli*. Infect Immun 80(7):2276–2285

108. Nemec KN et al (2010) A host-specific factor is necessary for efficient folding of the autotransporter plasmid-encoded toxin. Biochimie 92(2):171–177

109. Orth D et al (2010) EspP, a serine protease of enterohemorrhagic *Escherichia coli*, impairs complement activation by cleaving complement factors C3/C3b and C5. Infect Immun 78(10):4294–4301

110. van Diemen PM et al (2005) Identification of enterohemorrhagic *Escherichia coli* O26:H-genes required for intestinal colonization in calves. Infect Immun 73(3):1735–1743

111. Puttamreddy S, Cornick NA, Minion FC (2010) Genome-wide transposon mutagenesis reveals a role for pO157 genes in biofilm development in *Escherichia coli* O157:H7 EDL933. Infect Immun 78(6):2377–2384

112. Dziva F et al (2007) EspP, a Type V-secreted serine protease of enterohaemorrhagic *Escherichia coli* O157:H7, influences intestinal colonization of calves and adherence to bovine primary intestinal epithelial cells. FEMS Microbiol Lett 271(2):258–264

113. Sui BQ, Dutta PR, Nataro JP (2003) Intracellular expression of the plasmid-encoded toxin from enteroaggregative *Escherichia coli*. Infect Immun 71(9):5364–5370

114. Villaseca JM et al (2000) Pet toxin from enteroaggregative *Escherichia coli* produces cellular damage associated with fodrin disruption. Infect Immun 68(10):5920–5927

115. Cappello RE et al (2011) Effects of the plasmid-encoded toxin of enteroaggregative *Escherichia coli* on focal adhesion complexes. FEMS Immunol Med Microbiol 61(3):301–314

116. Navarro-Garcia F et al (2007) Intoxication of epithelial cells by plasmid-encoded toxin requires clathrin-mediated endocytosis. Microbiology 153(Pt 9):2828–2838

117. Navarro-Garcia F et al (2007) Pet, a non-AB toxin, is transported and translocated into epithelial cells by a retrograde trafficking pathway. Infect Immun 75(5):2101–2109

118. Vidal JE, Navarro-Garcia F (2008) EspC translocation into epithelial cells by enteropathogenic *Escherichia coli* requires a concerted participation of type V and III secretion systems. Cell Microbiol 10(10):1975–1986

119. Vidal JE, Navarro-Garcia F (2006) Efficient translocation of EspC into epithelial cells depends on enteropathogenic *Escherichia coli* and host cell contact. Infect Immun 74(4):2293–2303

120. Drago-Serrano ME, Parra SG, Manjarrez-Hernandez HA (2006) EspC, an autotransporter protein secreted by enteropathogenic *Escherichia coli* (EPEC), displays protease activity on human hemoglobin. FEMS Microbiol Lett 265(1):35–40

121. Kenny B, Finlay BB (1995) Protein secretion by enteropathogenic *Escherichia coli* is essential for transducing signals to epithelial cells. Proc Natl Acad Sci USA 92(17):7991–7995

122. Mundy R et al (2005) *Citrobacter rodentium* of mice and man. Cell Microbiol 7(12):1697–1706

123. Al-Hasani K et al (2001) Genetic organization of the she pathogenicity island in *Shigella flexneri 2a*. Microb Pathog 30(1):1–8

124. Kotloff KL et al (1999) Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ 77(8):651–666

125. Levine MM et al (2007) Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. Nat Rev Microbiol 5(7):540–553

126. Ashida H et al (2011) Shigella are versatile mucosal pathogens that circumvent the host innate immune system. Curr Opin Immunol 23(4):448–455

127. Niyogi SK, Vargas M, Vila J (2004) Prevalence of the sat, set and sen genes among diverse serotypes of *Shigella flexneri* strains isolated from patients with acute diarrhoea. Clin Microbiol Infect 10(6):574–576

128. Wei J et al (2003) Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. Infect Immun 71(5):2775–2786

129. Yang F et al (2005) Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res 33(19):6445–6458

130. Guyer DM et al (2002) Sat, the secreted autotransporter toxin of uropathogenic *Escherichia coli*, is a vacuolating cytotoxin for bladder and kidney epithelial cells. Infect Immun 70(8):4539–4546

131. Guignot J et al (2007) The secreted autotransporter toxin, Sat, functions as a virulence factor in Afa/Dr diffusely adhering *Escherichia coli* by promoting lesions in tight junction of polarized epithelial cells. Cell Microbiol 9(1):204–221

132. Lievin-Le Moal V et al (2011) Secreted autotransporter toxin (Sat) triggers autophagy in epithelial cells that relies on cell detachment. Cell Microbiol 13(7):992–1013

133. Yang Z, Klionsky DJ (2010) Mammalian autophagy: core molecular machinery and signaling regulation. Curr Opin Cell Biol 22(2):124–131

134. Kostakioti M, Stathopoulos C (2004) Functional analysis of the Tsh autotransporter from an avian pathogenic *Escherichia coli* strain. Infect Immun 72(10):5548–5554

135. Provence DL, Curtiss R 3rd (1994) Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect Immun 62(4):1369–1380

136. Benjelloun-Touimi Z et al (1998) SepA, the 110 kDa protein secreted by *Shigella flexneri*: two-domain structure and proteolytic activity. Microbiology 144(Pt 7):1815–1822

137. Otto BR et al (2002) *Escherichia coli* hemoglobin protease autotransporter contributes to synergistic abscess formation and heme-dependent growth of *Bacteroides fragilis*. Infect Immun 70(1):5–10

138. Dozois CM et al (2000) Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region. Infect Immun 68(7):4145–4154

139. Restieri C et al (2007) Autotransporter-encoding sequences are phylogenetically distributed among *Escherichia coli* clinical isolates and reference strains. Appl Environ Microbiol 73(5):1553–1562

140. Gomis SM et al (2001) Phenotypic and genotypic characterization of virulence factors of *Escherichia coli* isolated from broiler chickens with simultaneous occurrence of cellulitis and other colibacillosis lesions. Can J Vet Res 65(1):1–6

141. Navarro-Garcia F et al (2010) Pic, an autotransporter protein secreted by different pathogens in the Enterobacteriaceae family, is a potent mucus secretagogue. Infect Immun 78(10):4101–4109

142. Freeman GJ et al (2010) TIM genes: a family of cell surface phosphatidylserine receptors that regulate innate and adaptive immunity. Immunol Rev 235(1):172–189

143. Carlow DA et al (2009) PSGL-1 function in immunity and steady state homeostasis. Immunol Rev 230(1):75–96

144. Naor D, Sionov RV, Ish-Shalom D (1997) CD44: structure, function, and association with the malignant process. Adv Cancer Res 71:241–319

145. Fukuda M, Tsuboi S (1999) Mucin-type O-glycans and leukosialin. Biochim Biophys Acta 1455(2–3):205–217

146. Rosenstein Y, Santana A, Pedraza-Alva G (1999) CD43, a molecule with multiple functions. Immunol Res 20(2):89–99

147. Chen SC et al (2004) Cross-linking of P-selectin glycoprotein ligand-1 induces death of activated T cells. Blood 104(10):3233–3242

148. Artus C et al (2006) CD44 ligation induces caspase-independent cell death via a novel calpain/AIF pathway in human erythroleukemia cells. Oncogene 25(42):5741–5751

149. Bazil V et al (1995) Apoptosis of human hematopoietic progenitor cells induced by crosslinking of surface CD43, the major sialoglycoprotein of leukocytes. Blood 86(2):502–511

150. Klaus SJ, Sidorenko SP, Clark EA (1996) CD45 ligation induces programmed cell death in T and B lymphocytes. J Immunol 156(8):2743–2753

151. Kum WW et al (2010) Impaired innate immune response and enhanced pathology during *Citrobacter rodentium* infection in mice lacking functional P-selectin. Cell Microbiol 12(9):1250–1271

152. Grassl GA et al (2010) CD34 mediates intestinal inflammation in Salmonella-infected mice. Cell Microbiol 12(11):1562–1575

153. Niess JH et al (2005) CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. Science 307(5707):254–258

154. Kum WW et al (2009) Lack of functional P-selectin ligand exacerbates *Salmonella serovar typhimurium* infection. J Immunol 182(10):6550–6561

155. Medina-Contreras O et al (2011) CX3CR1 regulates intestinal macrophage homeostasis, bacterial translocation, and colitogenic Th17 responses in mice. J Clin Invest 121(12):4787–4795

156. Nataro JP (2011) Outbreak of hemolytic-uremic syndrome linked to Shiga toxin-producing enteroaggregative *Escherichia coli* O104:H4. Pediatr Res 70(3):221

157. Benjelloun-Touimi Z, Sansonetti PJ, Parsot C (1995) SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. Mol Microbiol 17(1):123–135

158. Boisen N et al (2012) Genomic characterization of enteroaggregative *Escherichia coli* from children in Mali. J Infect Dis 205(3):431–444

159. Coron E et al (2009) Characterisation of early mucosal and neuronal lesions following *Shigella flexneri* infection in human colon. PLoS ONE 4(3):e4713

160. Roy K et al (2011) Adhesin degradation accelerates delivery of heat-labile toxin by enterotoxigenic *Escherichia coli*. J Biol Chem 286(34):29771–29779

161. O'Ryan M, Prado V, Pickering LK (2005) A millennium update on pediatric diarrheal illness in the developing world. Semin Pediatr Infect Dis 16(2):125–136

162. Kiefer F, et al (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37:D387–392

163. Petersen TN et al (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786

164. Sievers F et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539

165. Dereeper A, et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469

166. Guindon S et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307–321

167. Chevenet F et al (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinf 7:439